THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Donald McGillivray

TITLE OF THESIS: Dependence of Resolution on Specimen Thickness in Thermionic Scanning Transmission Electron Microscopy.

DEGREE FOR WHICH THESIS WAS PRESENTED: M. Sc.

YEAR THIS DEGREE GRANTED: 1977

THE UNIVERSITY OF ALBERTA

DEPENDENCE OF RESOLUTION ON SPECIMEN THICKNESS IN THERMIONIC SCANNING TRANSMISSION ELECTRON MICROSCOPY

by

© DONALD McGILLIVRAY

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE

OF MASTER OF SCIENCE

DEPARTMENT OF PHYSICS

EDMONTON, ALBERTA

FALL, 1977

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research, for acceptance, a thesis entitled DEPENDENCE OF RESOLUTION ON SPECIMEN THICKNESS IN THERMIONIC SCANNING TRANSMISSION ELECTRON MICROSCOPY submitted by Donald McGillivray in partial fulfillment of the requirements for the degree of Master of Science.

ABSTRACT

A project was undertaken to study the thickness dependence of resolution in scanning transmission electron microscopy.

The standard expression for resolution in a scanning transmission electron microscope is considered in detail. Using known or measured parameters this expression is used to calculate the theoretical resolution obtained with thin specimens. The actual resolution was measured by spectrum analysis and compared to the theoretical value. Theory and experiment were found to be in good agreement where the demagnified Gaussian spot size is dominant but poor where spherical aberration is dominant. Possible explanations for this result are discussed.

Some preliminary measurements were also made on thick crystalline specimens using edge spreading as a measure of resolution. It would appear from these crude experiments that noise is always the limiting factor in scanning transmission electron microscopy. In these preliminary experiments conventional transmission electron microscopy gave better resolution than scanning.

ACKNOWLEDGEMENTS

I would like to acknowledge the valuable advice of Dr. P.S. Turner during the experimental and initial write-up phase of this project. I would also like to thank Dr. S.S. Sheinin for his advice while completing this thesis. Thanks are also due to Dr. J. Strauss and to Dr. G. Corbett for provision of specimens. Finally, I would like to thank J.C. Brunel for his maintenance of the microscope.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF PLATES

CHAPTER 1

THE DEVELOPMENT OF THE SCANNING TRANSMISSION ELECTRON MICROSCOPE

1.1 Introduction

The techniques of conventional transmission electron microscopy and scanning transmission electron microscopy both employ the transmitted electrons as a signal to form an electron microscope image. One potential advantage of scanning transmission electron microscopy over conventional transmission microscopy is the possibility of obtaining useful images from thicker specimens than can normally be imaged in conventional microscopy. This possibility arises because chromatic aberration is essentially absent under scanning conditions due to the fact that there are no imaging lenses past the specimen. The opportunity of investigating this possibility arose when in 1972, one of the first JEOL 100B transmission electron microscopes with scanning attachment was obtained by the Department of Physics. Therefore, a project was undertaken to study resolution in scanning transmission electron microscopy.

This thesis reports work whose initial object was to determine the relationship between image resolution and specimen thickness in scanning transmission microscopy. As a first step it was necessary to determine optimum operating conditions and the resolution available for thin specimens. Several methods of resolution measurement were explored before an adequate one was found. Some preliminary experiments were then carried out to determine the resolution available in thick crystalline specimens.

In this chapter, the history of the scanning transmission electron microscope is reviewed and the microscope is compared to the conventional instrument. The literature relevant to the project is reviewed and the aims of the project are outlined in detail. In chapter two, the microscope is described, values for the various microscope parameters are given and the methods whereby these were obtained are also given. The various methods of measuring resolution are then considered. Chapter three is concerned with the results of the resolution measurements, discussion of the results and suggestions for further work.

1.2 Electron Microscopy

The most common type of electron microscope is the conventional transmission electron microscope (CTEM) which

is optically very similar to the standard light microscope. It has an electron gun that serves as a source of illumination. The electrons are then directed onto the specimen by one or two magnetic lenses which act as a condenser. The specimen itself is placed in or above the field of the objective lens which provides the first stage of magnification. Subsequent stages of magnification are provided by one or two intermediate lenses and the projector lens. The final image can be observed by allowing the electrons to impinge on a fluorescent screen or a photographic plate.

Another approach to obtaining an image is possible. Instead of illuminating the total area to be observed this area can be scanned in raster fashion by a small probe. Various signals corresponding to the transmitted primary electrons, backscattered primary electrons, secondary electrons or x-rays can be detected. An image can be obtained by modulating the intensity of the spot in a cathode ray tube so that it corresponds to the intensity of the desired signal and allowing this spot to move synchronously with the probe. If the transmitted electrons are collected, the instrument becomes a scanning transmission electron microscope (STEM). When the secondary electrons are detected, the instrument is referred to simply as a scanning electron microscope (SEM). It is this mode of operation that is in fact

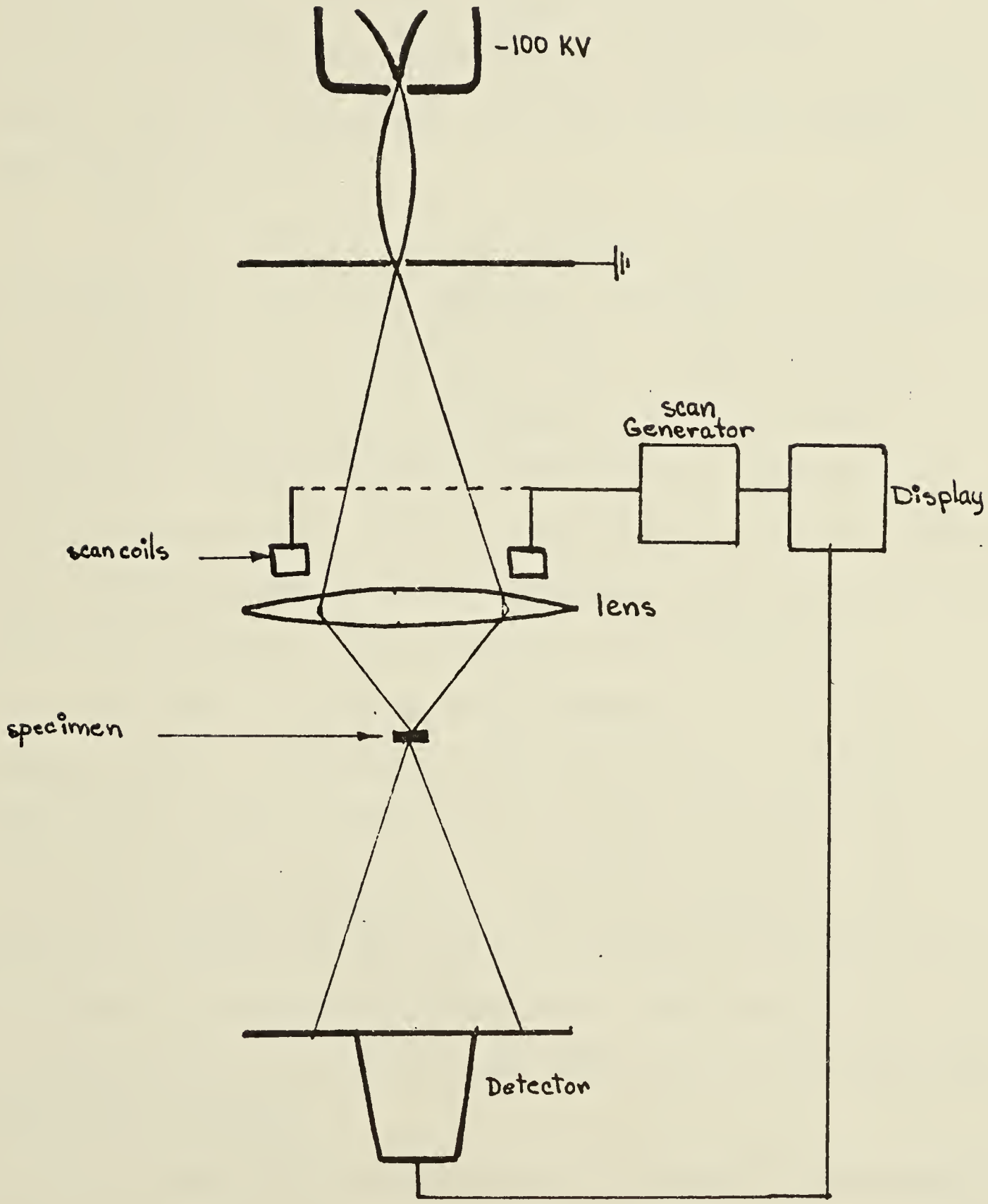
most widely employed.

Figure 1.1 shows a diagram of the basic STEM instrument. The source of electrons shown is a conventional electron gun. The beam is focussed to form a probe at the specimen by a series of lenses, represented in the diagram by a single lens. The scan coils just above this lens move the probe across the specimen in raster fashion. The transmitted electrons are detected by a scintillator-light pipe-photomultiplier arrangement and the resulting signal is amplified. This signal in turn determines the brightness of a spot on the display screen. This spot is driven synchronously with the scan coils by the scan generator.

1.3 Historical Development of STEM

The first suggestion of the possibility of a scanning microscope was made in 1935 by Knoll and the first such microscope was built in 1938 by von Ardenne (see Thorton, 1968). This instrument was a transmission type microscope where the intensity of the beam was recorded on a rotating photographic drum. Thus the first scanning microscope was a STEM. As the performance of STEM was at this time much worse than that of conventional instruments, work along these lines was not followed up. Von Ardenne however realized that resolution in STEM

Figure 1.1. Schematic diagram of STEM.



which depends on spot size need not be any worse than in CTEM since in principle a lens capable of imaging at a given resolution should be also capable of producing a demagnified electron probe of the same size from a suitable source. He also realized that there is no chromatic aberration in STEM because there are no imaging lenses beyond the specimen, so that STEM image should be free from this defect.

The initial work of von Ardenne stimulated the development of the first STEM by V. Zworykin, J. Hillier and R. Snyder (1942) (Thorton, 1968). This instrument used a cathode ray tube to display the picture and a detection and amplification system which consisted of a phosphor screen into which the electrons were accelerated. The resulting light was then converted into a current by a photomultiplier. They achieved 500 Å resolution.

In 1948, C. Oatley began research on the STEM (Oatley, 1972). The first one built by his group was completed in 1952 by McMullan. It employed a secondary emission electron multiplier, improved scanning and detection geometries and had two cathode ray tubes - one for viewing and one for photography. Over the years Oatley improved this microscope until in 1965 a commercial model was produced.

In 1966, A.V. Crewe became interested in the possibilities of obtaining high resolution in STEM. He realized

as von Ardenne had, that STEM should be able to achieve the same resolution as CTEM. He also realized that a major problem in attaining this resolution was that of obtaining an improved signal to noise ratio. (This problem will be discussed in more detail in this chapter and later in Appendix III.) In order to get an improved signal to noise ratio he developed a field emission gun which was able to achieve a brightness of about 1000 times that of a conventional gun while having a much smaller crossover. The small crossover allowed him to use a relatively simpler lens system and the high brightness allowed him to use small apertures to minimize the effect of spherical aberration until it was comparable to that of the diffraction effects. With this instrument, Crewe obtained 50 Å resolution in 1966 (Crewe, 1966, 1968). In 1969, he attained a resolution of approximately 5 Å at 20 kev which is comparable with conventional microscopy (Crewe, 1970).

These developments achieved by Oatley and Crewe led to an interest on the part of manufacturers of conventional microscopes in the development of scanning attachments for their instruments. In 1968, P.S. Ong presented a paper in which, starting from the idea of the x-ray microprobe, proposed an instrument capable of normal transmission microscopy, scanning microscopy and scanning transmission microscopy. He also presented a

few pictures that he had taken with a crude model. In 1970, H. Koike et.al. reported the addition of a scanning device to a JEM-100 type microscope and by 1972, such instruments were available commercially.

1.4 Comparison of CTEM and STEM

Two of the most important considerations in electron microscopy are resolution and image contrast. These two characteristics of CTEM and STEM are best compared through the principle of reciprocity. In this section resolution and contrast will first be considered for CTEM and the principle of reciprocity will then be discussed. A comparison of resolution and contrast in CTEM and STEM will then be carried out by means of reciprocity. Finally, some special characteristics not covered by reciprocity will be considered.

1.4.1 CTEM

a) Resolution

The spherical aberration of a conventional magnetic lens is always positive (see Appendix III) and thus this defect cannot be avoided in the electron microscope by using lens combinations. To find the resolution, therefore, both the disk of confusion due to the spherical aberration and the airy disk due to diffraction by

the lens aperture must be taken into account. It can then be shown (Hawkes, 1972) that when this is done the resolving power of a CTEM is

$$\rho = K\, C_s^{1/4}\, \lambda^{3/4}$$

where C_s is the coefficient of spherical aberration, λ is the wavelength of the electrons, K is a constant and ρ is the resolving power. The value of K to be used in the above expression is given by Haine (1961) to be 0.43. If we use this value and a spherical aberration coefficient of C_s = 1 mm, then ρ = 2 Å for 100 Kv electrons. This resolution has very nearly been attained in CTEM.

Another factor restricting resolving power in CTEM is chromatic aberration. This aberration arises from variations in electron energy or in the focussing fields. It is therefore influenced by the stability of high voltage supply and the lens currents. Chromatic aberration can also arise from inelastic scattering of the incident electrons in the specimen. The probability of inelastic scattering increases with specimen thickness. The degradation in resolution which occurs as a result of the energy spread of the transmitted electrons therefore also increases with specimen thickness. As a result chromatic aberration restricts CTEM specimens in practice to a few thousand angstroms thickness.

b) Contrast

There are two basic types of contrast in CTEM (see, for example, Hawkes, 1972). One of these is amplitude contrast which in turn is of two types: diffraction and mass thickness contrast.

In amplitude contrast, electrons scattered by the specimen by more than a chosen angle are intercepted by an aperture placed in the back focal plane of the objective lens. The greater the number of electrons scattered outside the aperture, the smaller will be the number of electrons contributing to the image. Image contrast arises from local variations in the structure of the specimen which result in a corresponding variation in the number of electrons scattered outside the aperture.

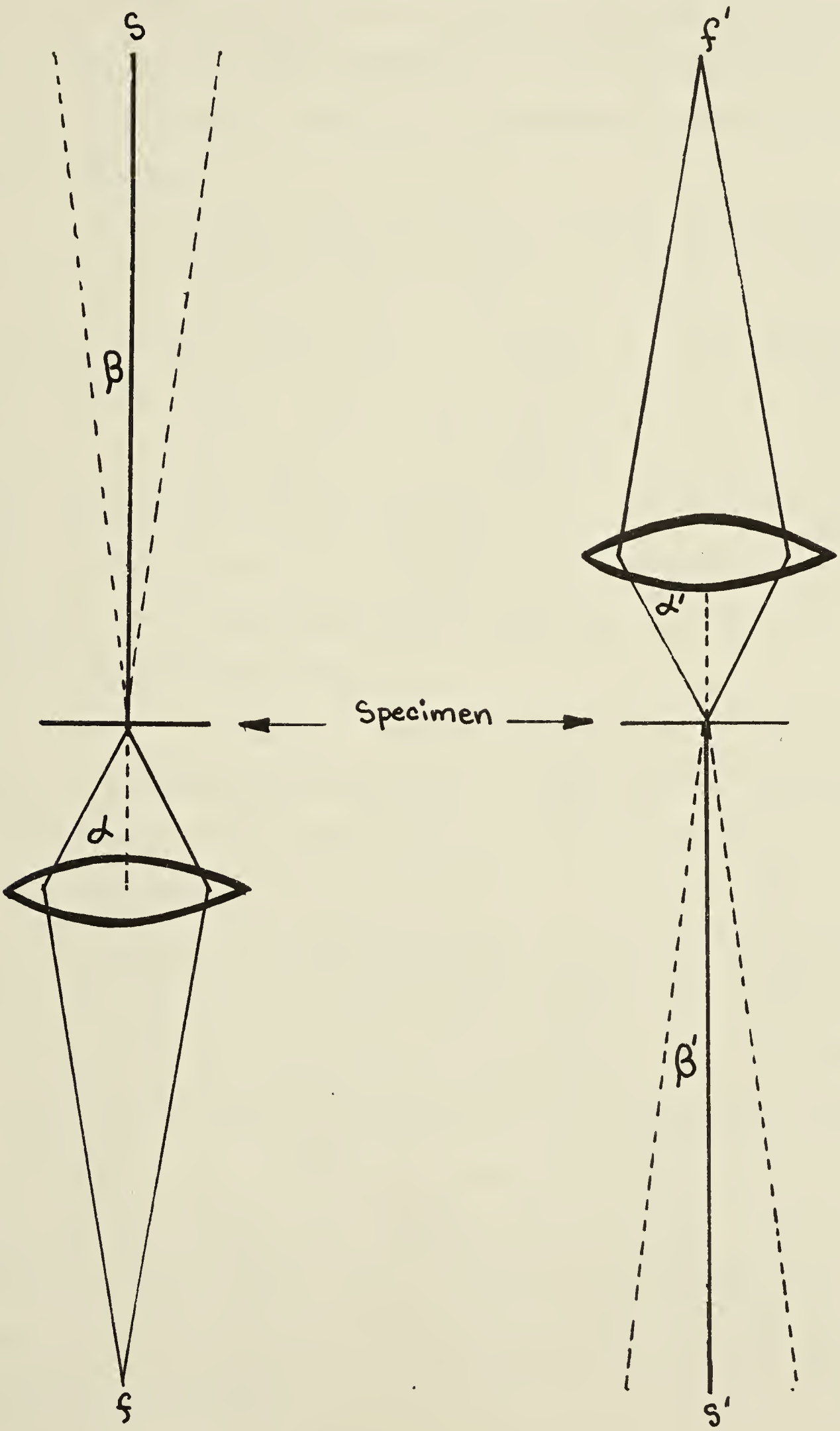
The other type of contrast is phase contrast. This type of contrast arises from a combination of the phase shift produced by the instrument through spherical aberration and objective lens defocus and the phase shift introduced by the specimen (Hawkes, 1972). The recent development of techniques in high resolution electron microscopy in imaging fine structure are based on the use of phase contrast.

1.4.2 STEM

a) Reciprocity

The early workers in STEM noticed that all the contrast mechanisms that are present in CTEM are also present in STEM (Crewe, 1968, 1970). This led J.M. Cowley (1969) to propose that all contrast in STEM could be explained by the principle of reciprocity. Later it was realized that this was also applicable to inelastically scattered electrons provided the energy loss does not significantly alter the wavelength or differential scattering cross section (Howie, 1972, Pogany and Turner, 1968). Let us now see how the principle of reciprocity explains the relationship between CTEM and STEM (see figure 1.2). The diagram on the left corresponds to the CTEM. Here we have a point source of electrons S some distance from the specimen. These electrons are scattered by the specimen and the electrons within the half angle α defined by an aperture are focussed onto a detector f (usually film). The diagram on the right is for STEM. Here the source of electrons f' is focussed to a point on the specimen with all the electrons in the half angle α' defined by an aperture. The scattered electrons are then detected by a point detector S'. If $\alpha = \alpha'$ then we see that STEM is identical to CTEM with reversed electron trajectories. The principle of reciprocity states that the intensity at

Figure 1.2. Comparison of STEM and CTEM.



f due to the source at S is identical to the intensity at S' due to the same source at f' i.e. interchanging source and detector make no difference to the intensity recorded.

In practice, of course, an electron source must be of finite size so that in CTEM we must specify the angle of illumination by a half angle β. Also detectors must have a finite size so that in STEM we must specify an angle of detection β'. Even here it is easy to see that the same result holds if $\beta = \beta'$. The fact that the STEM beam is scanned over the specimen is equivalent to scanning the CTEM image with a detector in order to record it (Cowley, 1969).

The principle of reciprocity can therefore be used to determine the contrast in STEM from a knowledge of the contrast in CTEM provided that the angles α and β shown in figure 1.3 are equal to α' and β'. This result is also useful in determining the effect on the resolution of variations in electron optical parameters. According to Cowley (1969), increasing the detector size in STEM is thus equivalent to increasing the illumination angle in CTEM and thus results in poorer resolution just as increasing illumination defocus does in CTEM. Similarly increasing the source size in STEM is equivalent to increasing the size of f in CTEM and thus results in a loss of resolution.

b) Aspects of STEM Not Covered by Reciprocity

In spite of equivalence by reciprocity, there are several potential advantages which STEM might have over CTEM. These will be discussed in this section.

i) Observation of Thick Specimen

One of the basic problems in electron microscopy is relating the structure of thin films to bulk materials. In order that this can more adequately be done it is desirable to have specimens which are as thick as possible. The desire to observe thick specimens, for example, has stimulated the development of high voltage electron microscopy. STEM also has the potential advantage of being able to observe thick specimens. This arises from the fact that in STEM, electrons are not imaged past the specimen and thus the chromatic aberration due to the energy loss in the specimen, which occurs in CTEM, is avoided. However, observation of thick specimens is then limited by the fact that scattering in the specimen will lead to beam broadening. This results in decreased resolution as explained more fully in the following paragraph.

Once the electrons enter the specimen, they suffer a series of random collisions. These collisions are of two types - elastic and inelastic. In the elastic case the incoming electron interacts with the

nucleus. Because the nucleus is so much heavier than the electron, a negligible amount of energy is lost by the electrons and the direction of the electron is changed. On the other hand if the incoming electron collides with another electron, then a significant amount of energy is likely to be lost because the masses are the same. We are however only interested in the angular distribution of the scattered electrons because there is no focussing of the electrons after the specimen and consequently energy losses are not important. Generally the elastic events involve much larger scattering angles than the inelastic scattering events. This results in an effective broadening of the beam as the electrons move through the specimen. The broadening will, of course, become more pronounced as specimen thickness increases. Notice that it is the beam broadening at the depth of the feature of interest within the specimen that affects the resolution. Thus resolution at the top of a specimen will be unaffected but the examination of a feature well inside the specimen can be significantly affected by broadening of the beam. This is known as the Hashimoto top-bottom effect (Reimer, 1972). However with increasing thickness, the number of electrons that are scattered outside the detector angle increases. This results in a decrease in the signal to noise ratio and it may become necessary to change the operating conditions

of the instrument in order to reduce the statistical noise.

ii) Other Advantages of STEM

Another advantage of STEM is that it is possible to carry out electronic signal processing of the image directly. This is usually done by means of variable amplification and a bias voltage which enables the elimination of the background signal. It is also possible to carry out such a procedure on CTEM micrograph but the procedure is much more involved. In addition it is easy to carry out energy analysis in STEM. Here only those electrons of a certain energy are admitted to the detector. Different chemical elements have different characteristic energy losses associated with them, so this technique can be used to determine the location of different elements in the specimen. This can also be done in CTEM but is made more difficult because of the need for imaging by lenses past the analyser. However, both these techniques suffer from the disadvantage that they result in a decrease in the effective signal to noise ratio. Any attempt to achieve better resolution results in the noise quickly becoming intolerable. Much longer exposure times would overcome the noise problem but reduce its practical usefulness. In any case machine drift with respect to time would probably preclude them.

c) Thermionic STEM

Resolution in STEM is determined by the size of the spot incident on the specimen. In order to obtain high resolution the spot should be as small as possible. In this section, some of the factors affecting the spot size obtained in a microscope which uses a thermionic electron gun are discussed.

The first general consideration to note is that in determining the minimum spot size attainable lens aberrations result in a point object being imaged as a disc of diameter d_{eff} which depends on the half angle of illumination α. These effects can be summarized as follows:

1) Spherical aberration $d_s = \frac{1}{2}C_s\alpha^3$ where C_s is the coefficient of spherical aberration.
2) Chromatic aberration $d_c = C_c \frac{\Delta V}{V} \alpha$ where C_c is the coefficient of chromatic aberration and ΔV is the change from the voltage V that the average electron has.
3) Airy disk $d_d = \frac{1.22\lambda}{\alpha}$ where λ is the electron wavelength.

The expressions in 1) and 2) are justified in Appendix III when lens aberrations are considered and 3) is just the standard Airy disk diameter due to diffraction. The spot size that would have been obtained if no

aberrations were present is called the Gaussian spot and is of size d_g. This is determined entirely by the focal lengths of the lenses and the lens geometry. The aberration disks should be convoluted in order to obtain d_{eff}. However, if all the error disks are Gaussian, then the result is the same as addition in quadrature. Usually the Gaussian form is just assumed without being stated (e.g. Riemer, 1972) and the expression for d_{eff} becomes

$$d_{eff}^2 = d_g^2 + (\tfrac{1}{2}C_s\alpha^3)^2 + (C_c \frac{\Delta V}{V} \alpha)^2 + (1.22 \frac{\lambda}{\alpha})^2 \quad . \quad (1.1)$$

Equation (1.1) shows that there is an optimum value of α for which a minimum d_{eff} will be obtained. For a fixed pre-specimen aperture d_g is a function of α. However, it is possible at least in principle to set the demagnification of the gun spot so high that d_g becomes an insignificant contribution to d_{eff} and then set the angle α to the optimum value by means of the appropriate aperture. In order for this to occur d_g would have to be of the order of 1 Å and the optimum α would be about 10^{-2} radians. This is fine for STEM with a field emission gun but thermionic STEM suffers from a shortage of electrons. A hot tungsten filament whose emitted electrons are subsequently accelerated to 100 kv has a maximum brightness of 3×10^5 amps/cm^2 steradian

(see Appendix III for a complete discussion of this point). If we use these values then the total current in the probe is about 7.4×10^{-15} amps or about 1.18×10^{5} electrons per second. If the exposure time is 50 seconds and there are 10^{6} picture points in a frame, then each point will receive just six electrons in the signal resulting in a shot noise in the picture, due to the noise inherent in the probe, of 40%.

Because of this high noise level it is necessary to increase the Gaussian spot size (increasing the half angle α substantially will result in intolerable values of spherical aberration). If the Gaussian spot is increased to 30 Å and the half angle α is 10^{-2} radians, then the probe receives 1.2×10^{8} electrons/sec. Using the same assumptions as before this will result in a "shot noise current" of 1.3%. Longer exposure time would in principle reduce the need for such large Gaussian spot sizes but is in practice impractical since the electronics are subject to drift over long periods of time. In any event long exposure time would make the machine of dubious practical value. It is therefore obvious that the first two terms in equation (1.1) completely dominate in thermionic STEM while the last two terms must be taken into account only when a field emission gun is used.

d) Early Work on STEM Using Thermionic Sources

It is evident from the effect of beam broadening discussed in Section 1.4 c) that the resolution attainable in STEM will decrease with increasing specimen thickness. One of the first attempts to measure the relationship between resolution and thickness was made by Takashima, Hashimoto and Kimoto (1970). For their specimen they used a wedge of glass where the angle of the wedge was 90°. They coated part of the wedge with a film of Ag about 0.2 μ thick in order to produce a sharp step. They measured the rise distance for the signal to go from 10% to 90% of full scale. They found that the resolution d is given by

$$d = d(0) + kt$$

where d(0) is the probe size, t is the thickness and k is a factor that depends on the accelerating voltage which in principle should be equal to the half angle of the probe but in practice was much bigger. One of the weaknesses of their work is that they apparently did not consider the top-bottom effect.

Because STEM is a relatively new field there is little quantitative information on the relationship between specimen thickness and resolution. Perhaps the key paper from a theoretical point of view is the one due to Reimer in 1972 who calculated $d_B(t)$ for amorphous

carbon ($d_B(t)$ is the diameter that a point at the surface is broadened to at depth t). He pointed out that the theory of multiple scattering cannot be used to predict the effective local distribution at the bottom of the specimen if the electrons which reach the detector are determined by an aperture. To overcome this Monte Carlo calculations were carried out. However, this method cannot be used for small limiting apertures and large thicknesses because so few electrons go through the aperture that the computation time increases unreasonably. In order to carry out this calculation, Reimer defined a set of reduced coordinates such that the film thickness was expressed in terms of the number of mean free paths between elastic collisions (a fixed ratio of inelastic to elastic collisions was also assumed but this ratio did not substantially affect Reimer's results). The broadening obtained from the calculation was then plotted as a function of a reduced angle equal to the aperture angle divided by the angle at which the scattered amplitude is half maximum. This effectively put the energy dependence of the broadening in term of parameters that did not directly enter the calculations and thus made the calculations independent of energy.

In addition, there is another effect which tends to degrade the available resolution in a thick specimen. If the microscope is focussed on a feature at depth t_o

and the beam is also incident on a feature at depth t', then a point at t_o will become a disk at t' (see Oatley, 1972). We will call this disk d_f. If α_o is the angle formed by the extreme rays then

$$d_f = 2\alpha_o|t_o - t'| \quad .$$

In general d_f can be made zero by choosing the right depth of focus in order to see the feature of interest. Reimer only considers this effect for $t_o = 0$.

As mentioned previously, contributions to an error disk in electron optics are normally added in quadrature to obtain the size of the disk. Reimer combines these error disks ($d_B(t)$ and d_f) with the probe size for thin specimens obtaining

$$d^2 = d_{eff}^2 + d_B^2(t) + d_f^2$$

where d is the size of the error disk at the feature of interest.

In order to be visible in STEM a feature must not only be bigger than d but must have sufficient contrast to be seen above the noise. Reimer therefore considers how large a thickness variation must be in order to be seen. It is possible to show that there is a physical limit to the electron optical brightness B' of a filament where B' is the current of electrons in a crossover per unit area per unit solid angle (see Appendix III

for more on this point). If the probe was an unaberrated spot its area would be $\pi d_g^2/4$. Then the total current of the probe is

$$I = \frac{\pi^2 B' \alpha^2 d_g^2}{4} \quad .$$

If we have n^2 points in a picture and it takes τ seconds to take a picture, then the number of electrons N per picture point is

$$N = \frac{I\tau}{n^2 e} \quad .$$

This gives us a signal to noise ratio just due to the fluctuations in the electron current of

$$\frac{S}{N} = \frac{N}{N^{\frac{1}{2}}} = \frac{I^{\frac{1}{2}}\tau^{\frac{1}{2}}}{ne^{\frac{1}{2}}} \quad .$$

Then at depth t we get a signal to noise ratio of

$$\frac{S}{N} = \frac{I(t) - I(t+\Delta t)}{I_N(t)}$$

for a variation of thickness of Δt in the specimen, where $I_N(t)$ is the noise current at depth t ($I_N(t) = I(t)^{\frac{1}{2}}$).

Let the transmission function be defined as

$$T(t) = \frac{I(t)}{I_o}$$

where I_o is the current at the surface, then

$$\frac{S}{N} = \frac{\frac{I(t)}{I} - \frac{I(t+\Delta t)}{I}}{\frac{I_N(t)}{I_N}} \times \frac{I_N}{I} = \frac{T(t) - T(t+\Delta t)}{(T(t))^{\frac{1}{2}}} \frac{I_N}{N}$$

$$= \frac{dT}{dt} \Delta t \; T^{-\frac{1}{2}} \left(\frac{ne}{\tau^{\frac{1}{2}} I^{\frac{1}{2}}}\right) \quad .$$

If we know the acceptable signal to noise ratio and the transmission function T(t), then we can predict whether a certain thickness variation will be visible.

Gentch and Reimer (1972) and Gentch, Gilde and Reimer (1973) carried out some experiments designed to check the theory described above. They deposited a thin indium layer on a carbon coated formvar film. This resulted in small flat indium islands with sharp edges. Then the film was coated on one side with polystyrene spheres of different sizes. They took pictures with a thermionic gun in both CTEM and STEM modes for either side of the film up. The top-bottom effect was clearly observed. They took densitometer measurements of the edge of the indium crystals, where the beam had passed through the spheres first, in order to obtain a thickness vs. resolution curve. (This procedure is difficult as absolute densitometry is involved. In addition, the transfer function from electron intensity to photographic density can be non-linear unless great care is taken. Unless the exposure is kept constant or variable exposure

is compensated for, an inconsistent factor will be introduced into the results. In addition, the density exposure curve is very dependent on constant development conditions. The authors only state that photodensitometer traces were taken of the negatives.) Using data obtained in this way, they calculated the beam broadening and compared the results with Reimer's earlier calculations. The agreement was very poor because there was no dependence in the experimental results on the objective aperture as opposed to Reimer's prediction of strong dependence. They concluded that this was because of the large angle of illumination and the fact that the objective aperture in STEM was not in the back focal plane.

The other major theoretical attempt to define the relationship between resolution and thickness was by Sellar and Cowley in 1973. These authors made an approximation to describe the scattering for small α in order to be able to avoid Reimer's problem. They define the visibility as

$$V = \frac{I(obs)_{max} - I(obs)_{min}}{I(obs)_{max} + I(obs)_{min}}$$

and then considered three cases:

a) a layer of amorphous carbon

b) a layer of "biological material" (really just carbon that was half as dense)

c) an environmental cell.

Some of their calculations were applicable for 1 Mev electrons and some for 0.1 Mev electrons. They found that STEM had greater penetrating power and higher visibility at both voltages for cases a) and b) while case c) was rather unclear at high voltages. They appear, however, to be assuming that a field emission gun is used (they assume as many electrons as you need are available) and so the applicability of their results to the thermionic gun situation is doubtful.

1.5 The Project

Aside from these papers, there is a paucity of experimental results in the literature except for a number of pictures that qualitatively compare STEM and CTEM. Therefore the decision was taken to investigate the relationship between resolution and thickness. However, it was first necessary to determine the best operating conditions for the microscope.

The two important parameters in STEM are the aperture diameter above the specimen and the demagnification of the lenses. If the aperture diameter is known, then these can be reduced to one - the incident half angle (this is explained more fully in Section 3.1). This was measured by the convergent beam technique. At the same time the crossover spot size was measured

in order to be able to calculate d_g. Using data supplied by the manufacturer and making certain reasonable assumptions the theoretical resolution as a function of half angle was calculated.

An attempt was made to measure this function experimentally. The main difficulty here was determining the best method of measuring resolution. Those methods tried included visual scan to determine point to point resolution, diffractometer measurements and rise distance measurements. All these failed for various reasons. Finally, spectrum analysis of the detected signal was tried and proved to be successful. In general the results of this measurement agreed with the theoretical curve if the dominating factor determining resolution was the demagnified Gaussian spot size but disagreed if the spherical aberration dominated.

Reimer's work on amorphous carbon had come out by this time. In view of this and the limited time available it was decided to carry out some preliminary experiments on a crystalline specimen. An experiment similar to that of Gentch et.al. (1973) was carried out to the stage where initial results were obtained. These initial results agreed fairly well with what was theoretically expected.

CHAPTER II

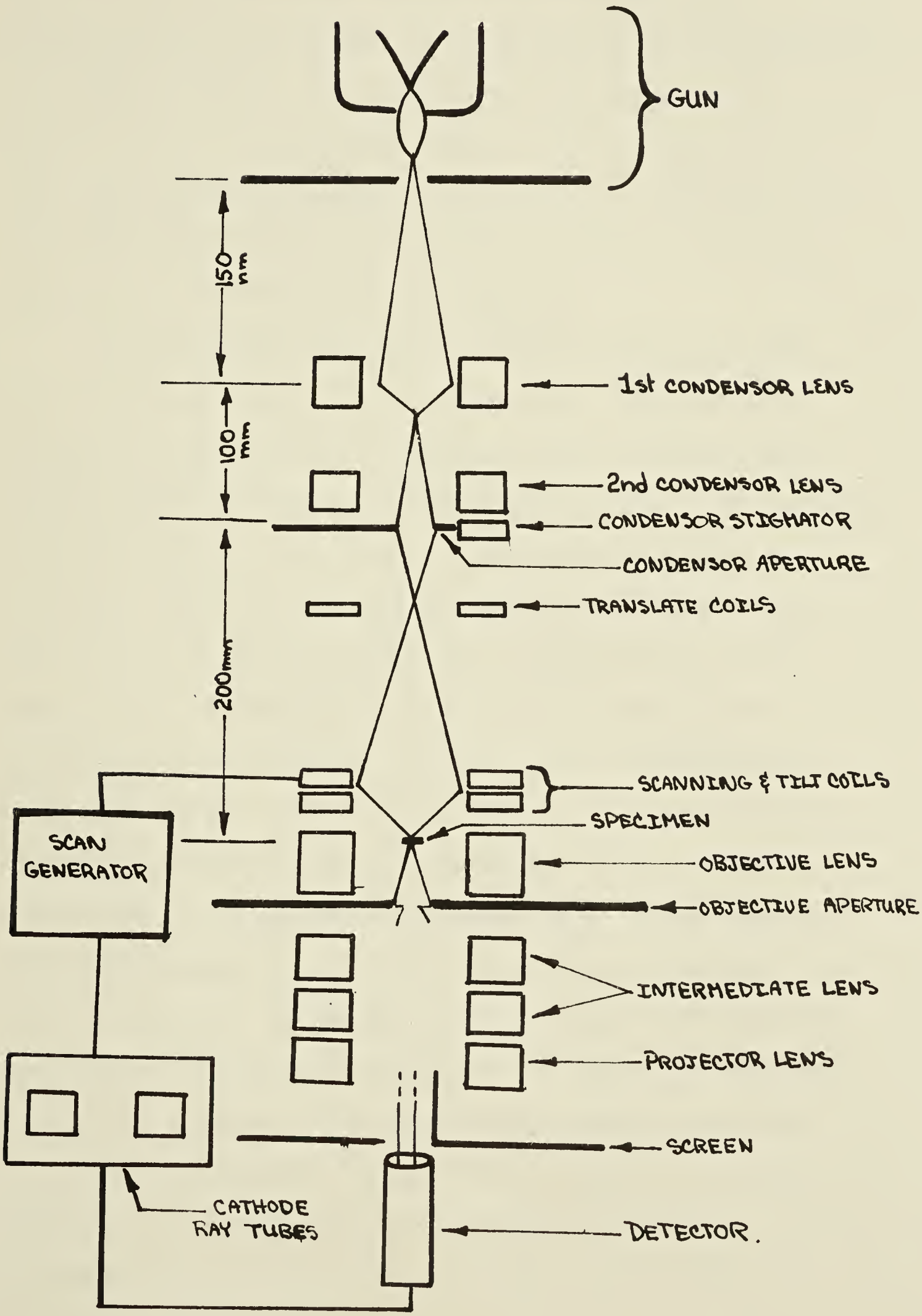
THE INSTRUMENT AND EXPERIMENTAL TECHNIQUE

In this chapter, the instrument used in carrying out the experiments presented in this thesis is described. The method whereby the various microscope parameters were determined is presented and the manner in which these parameters were used to compute the theoretical resolution is discussed. Finally various methods of measuring resolution are discussed.

2.1 The Instrument

During this project a JEOL 100B electron microscope with a scanning attachment was used to carry out the experimental work. Figure 2.1 is a schematic diagram of the microscope. In order to be able to convert the instrument from a CTEM to a STEM, the manufacturer has added a set of scanning coils just above the objective lens. These coils are activated by a set of scan generators which also drive the spot across the screens of the cathode ray tubes which are used for observation and photography. The intensity of these spots on the cathode ray tubes depends on the amplified signal from the electron detector (which

Figure 2.1. The microscope



consists of the standard scintillator-light pipe-photomultiplier arrangement) modified by the application of a bias voltage. This results in an image being formed on both cathode ray tubes. One of these has a long persistance phosphor and is used for viewing while the other has a short persistance phosphor to prevent halation and is used for photography.

As mentioned previously, high resolution STEM requires as small a probe as possible. Because the crossover produced by a thermionic gun is relatively large, the illumination system must provide a substantial demagnification. In order to achieve this, the first condenser lens is operated at maximum excitation thus decreasing its focal length and increasing the demagnification. In addition the objective lens current is increased so that the objective lens crossover is at the specimen. (The objective lens thus acts as an additional condenser lens, providing further demagnification.) The second condenser lens is now used to vary the demagnification and the condenser aperture is used to restrict the objective half angle thus reducing the effect of spherical aberration. Below the specimen either the objective or intermediate aperture is used to select the desired portion of the beam. (In STEM contrast is usually created by eliminating the scattered electrons so that the image corresponds to

the number of electrons unscattered by the specimen. However sometimes the scattered electrons or a portion thereof may be used to form the image and the main beam blocked out of the image. In the case of a crystalline specimen this corresponds to choosing a particular diffracted beam.) The only function of the intermediate lens is to focus the beam onto the detector.

Most of the physical data about the microscope was provided by the manufacturer. According to this data the distance from the gun crossover to the first condenser pole piece is 150 mm , the distance from the first condenser pole piece to the second is 100 mm and from the second condenser pole piece to the objective pole piece is 200 mm. In order to use these figures it was assumed that the principal planes corresponded with the centers of the lenses. These figures have been marked on Figure 3.1. According to the manufacturer, in the scanning mode and at 100 kv , the focal length of the first condenser lens is 1.4 mm and the focal length of the objective lens is 3.5 mm. For the objective lens, the coefficient of spherical aberration is 2.0 mm and the coefficient of chromatic aberration is 2.5 mm It can be shown that the aberrations of the other lenses do not contribute significantly to the spot size. Their values, therefore, are not listed below in Table I with the rest of the lens data.

TABLE I

lens/parameter	focal length	C_s	C_c
1st condenser	1.4 mm	-	-
2nd condenser	variable	-	-
objective	3.5 mm	2.0 mm	2.5 mm

2.2 Measurement of Other Parameters

i) The Half Angles

One of the most important parameters in determining the performance of a STEM is the half angle α of the incident beam at the specimen. This determines the size of the aberrations and the demagnified Gaussian spot size d_g. The half angle α, which depends on the demagnification, was measured using the convergent beam technique. This involved operating the illumination system in such a manner that a crossover occurred near the specimen. This results in a projection of the illuminated area of the specimen appearing on the CTEM viewing screem. (In this technique the crossover acts as a point source producing a cone of illumination. If the specimen is a metal, Bragg reflection will occur and multiple cones will appear.) The conditions used in these convergent beam technique measurements were identical to scanning conditions except that the

projector lens was on. Normally the scanning coils would be turned off. In the experiments, however, the scanning coils were left on since otherwise contamination built up rapidly on the speciment due to the passage of a large number of electrons through a small area. By using a high scanning magnification it was possible to have the probe move only very slightly and thus not significantly affect the position or shape of the resultant disks.

In order to measure α, a Mo specimen was used with the microscope set up for the convergent beam technique as described in the previous paragraph. Plate I is an example of the results obtained. In this case, as in all the others used to make the measurements, the 110 reflections were strongly excited and the Bragg angle was therefore 0.00831 radians.

The theory behind the convergent beam technique can be understood by referring to Figure 2.2. The beam is arranged so that a cone of illumination with half angle α has a crossover near the specimen. There it is Bragg diffracted and several cones of electrons with half angles of α emerge from the specimen at angles of 2β with respect to the directly transmitted beam where β is the Bragg angle. There is an effective length L to the recording surface (film). The size of the disks d on the surface is $2\alpha L$ and the

Figure 2.2. Convergent beam situation.

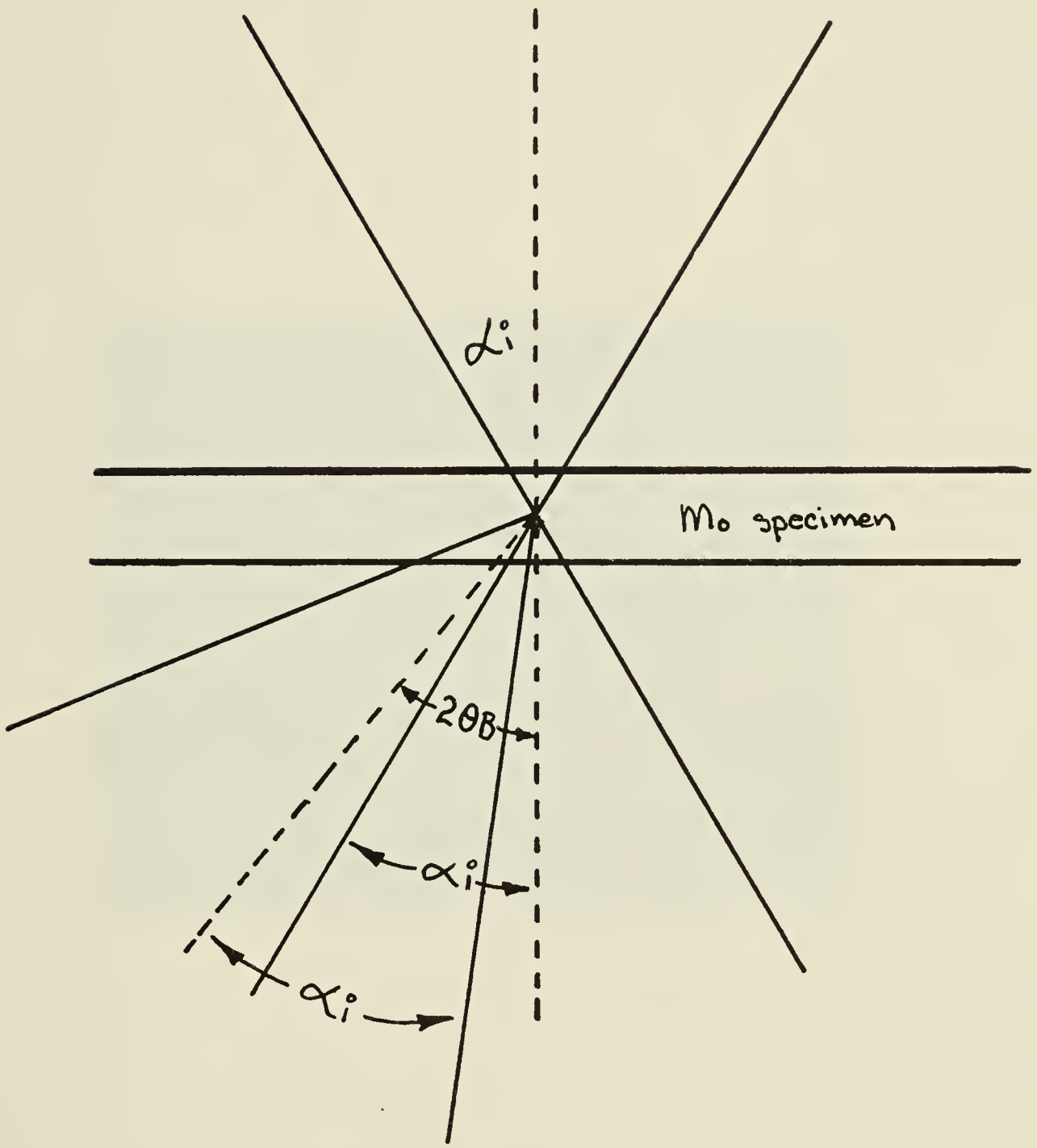
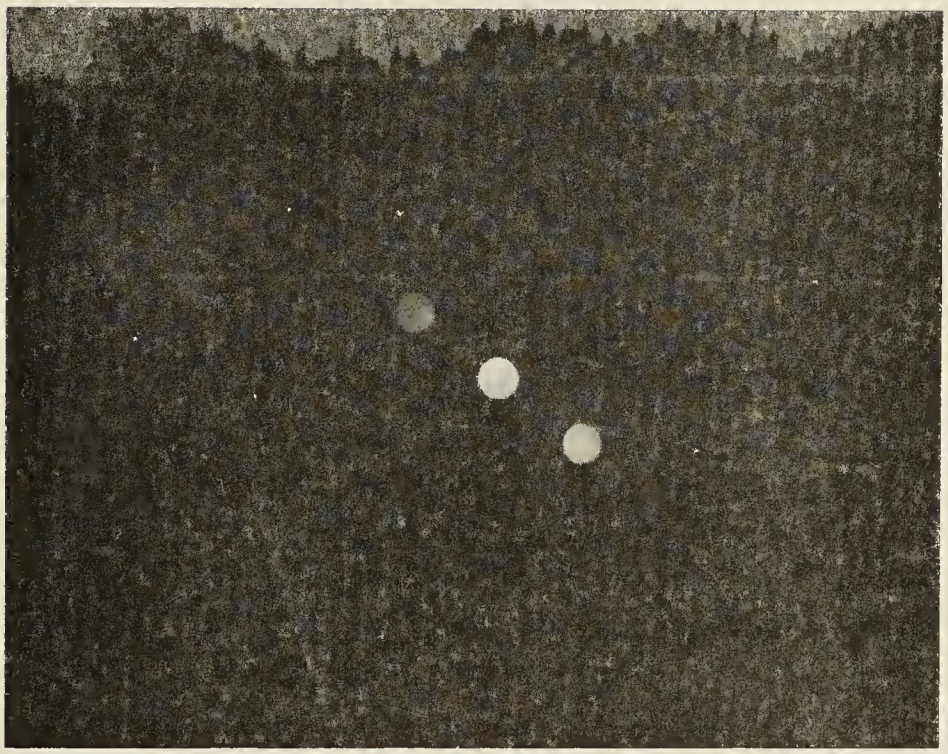


Plate I. Typical convergent beam pattern.



distance D between similiar edges of the disks is $2\beta L$. Thus it is possible to write

$$d/D = (2\alpha L)/(2\beta L) = \alpha/\beta$$

or

$$\alpha = (d/D)\beta \quad .$$

The half angle at the specimen can therefore be obtained by measuring d and D. Therefore, convergent beam patterns were taken for all possible second condenser lens settings and the distances d and D were measured.

In order to maximize the accuracy of the results it was convenient to express α as a function of current and to calculate a regression curve. It turned out to be possible to do this as will be seen below. As mentioned previously, the second condenser lens is used in STEM to vary the demagnification of the gun crossover. This lens is a weak lens. Using the fact that the magnetic field depends linearly on the current, it can be shown that the focal length of a weak lens, such as the second condenser lens, is given by

$$\frac{1}{f} = \frac{e}{8mV_r} \int_{-\infty}^{\infty} B^2 d_z \propto I^2$$

(El-Kareh and El-Kareh, 1970). Since the lens is weak it is reasonable to assume that the thin lens formula applies. If p is the image distance and q is the object

distance, rearranging the thin lens formula gives

$$q/p = q/f-1 \quad .$$

It is easy to show that q/p is the demagnification due to the second condenser lens. If it is assumed that the objective-object distance is not substantially affected by the second condenser image distance, then the half angle α is proportional to the demagnification of the second condenser lens. Using the previous relationship between current and focal length and the above form of the thin lens formula and using the fact that q (the second condenser object distance) is fixed by the setting of the first condenser lens, it is possible to show that

$$\alpha = aI^2 + b$$

where a and b are constants that depend on the lens parameters (focal length, principle planes, etc.). As it turns out the relationship is not strictly linear due to the fact that the object distance for the objective lens depends on the position of the second condenser crossover, contrary to what was assumed above. However, the effect on the objective lens demagnification is not very great so the above relationship is a good guide as to what to expect from a regression line.

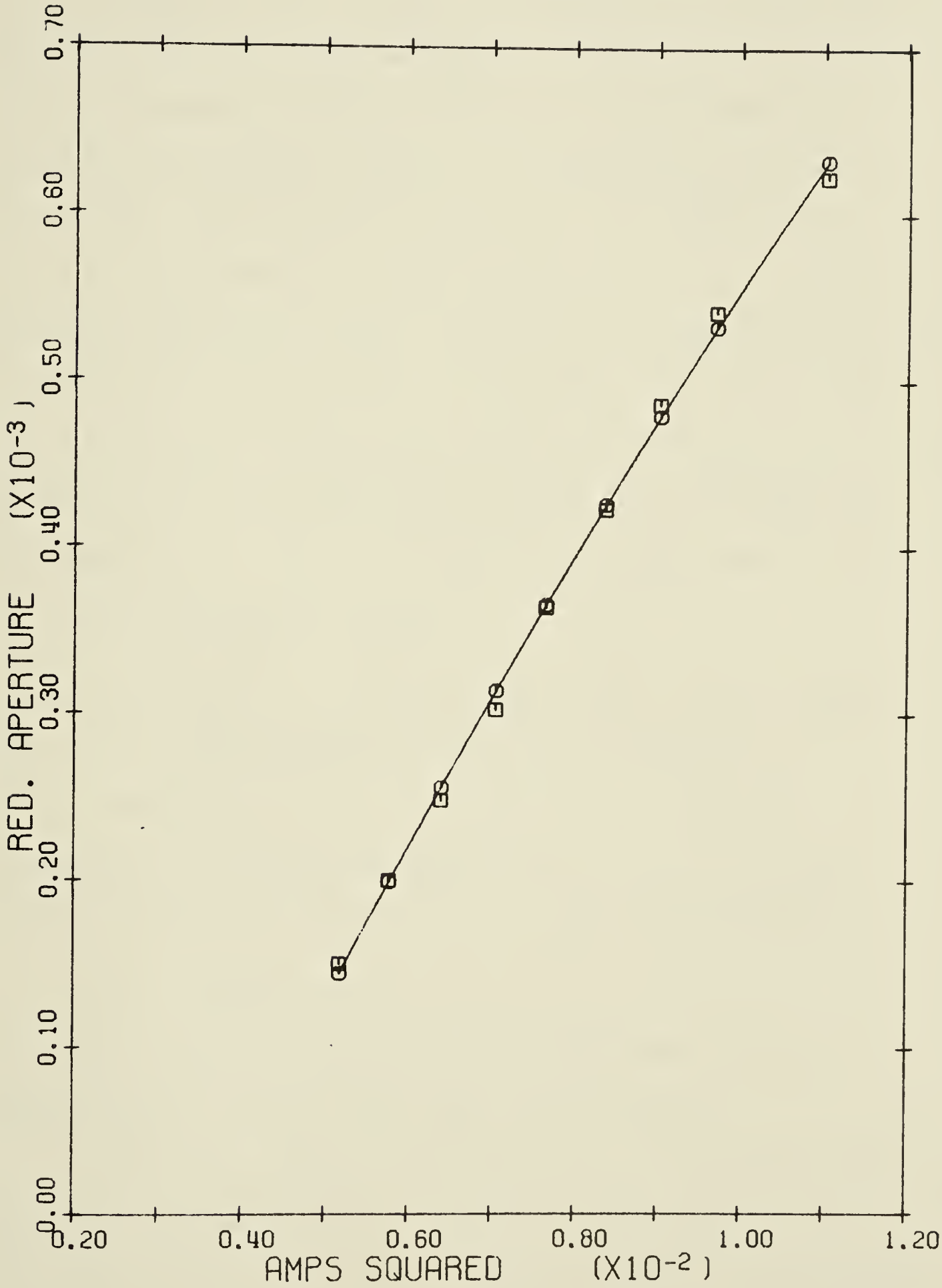
The convergent beam patterns were taken using only the smallest available aperature. However, if the

experiment is repeated, the results with different apertures must be proportional to the ratio of the aperture diameters. The size of the aperture used was measured using an optical microscope. In order to make the results universally applicable to all the apertures the half angles have been expressed in terms of reduced half angles (half angle divided by aperture diameter). The reduced half angle as a function of current squared is plotted in Figure 2.3 along with a least squares fit. The points indicated by the squares are the data points and the circles are the fitted points.

ii) The Gaussian Spot Size

The Gaussian spot size was also determined experimentally. The microscope was set up in the normal CTEM mode and the illumination spot was focussed down to its smallest size on the viewing screen while the microscope itself was focussed on a specimen. A micrograph was taken at low exposure and a densitometer trace taken across the spot. The procedure used to determine the size of the spot from this trace is given below. It should be noted that this experiment was not significantly affected by the aberrations since at every crossover the size of the aberration disks was much smaller than the size of the demagnified Gaussian spot size.

Figure 2.3. Calibration of the half-angles.



REDUCED HALF-ANGLE VS. C2 LENS CURRENT SQUARED

The size of the Gaussian spot was considered to be full width at half height. In order to be able to determine this, it was necessary to know the transfer function between incident electrons and photographic density on the plate. Density is defined in terms of scattering. If T is the proportion of a light beam that will be transmitted unscattered by the developed film then the photographic density D is proportional to log(1/T). According to Valentine (1966), the photographic density of film has a proportional relationship with exposure E if the illumination source is electrons and the exposure is low. Writing this and rearranging gives

$$E \propto \log(1/T) \propto - \log T \quad .$$

If s is the amount of light scattered out of the light beam due to the exposure of the film, then $T = 1 - s$. If the exposure is small, as was the case with the pictures of the spot described above, then s is small and it is possible to write, to a good approximation,

$$E \propto - \log(1-s) \propto s \quad .$$

This means that the exposure is proportional to the scattering out of the beam for low exposures.

Using the relationship derived in the last paragraph and the densitometer traces it was possible to find the size of the spot on the micrograph. The widths at half height of the two micrographs taken were found

by this method to be 13.6 mm and 15.6 mm. This gives an average measured spot size on the film surface of 14.6 mm. It is now necessary to relate this image of the gun cross-over to the actual crossover size.

Because the specimen used was a magnification grating, it is possible to find the size of the spot at the specimen. The magnification was calibrated by means of the grating to be 68,000 times, making the size of the image of the Gaussian spot at the specimen 0.215 u.

Now it is necessary to calculate the magnifications between the gun and the specimen in order to find the Gaussian spot size. In order to calculate the size of the first condenser crossover the magnification factor between it and the specimen crossover must be known. Using the thin lens formula and the data in Section 2.1, the nominal magnification is

$$M = (200/98.6) = 2.03 \quad .$$

However, the effect of the objective prefield must also be taken into account. In order to determine this, a diffraction pattern of Mo was taken using a second condenser aperture of 116 u. The spots in the pattern had a diameter of 1.2 mm. The pattern was indexed and the distance between the 000 and 002 spots found to be 18.0 mm. (The 002 Bragg angle is 0.01176 radians at 100 kev). Using the same reasoning as in the convergent

beam situation, the half angle γ_1 is found to be

$$\gamma_1 = (d/D)\gamma_2 = (1.2/18.0)0.01176 = 0.00078 \text{ radians.}$$

A 116 u. aperture was used at the second condenser lens. It was assumed that the condenser aperture was in the principle plane of the second condenser lens although this was not strictly true and introduces a small error. Ignoring the objective prefield and using the data from Section 3.1 and thin lens optics, the half angle would be

$$\gamma_1 = ((116 \text{ u.})/2)/(200 \text{ mm }) = 0.00029 \text{ radians} \quad .$$

Therefore the prefield magnification is

$$M_p = (0.00029/0.00078) = 0.372$$

and the size of the first condenser crossover is therefore

$$d_1 = ((0.215/(200/98.6))(1/0.372) = 0.285 \text{ u.}$$

Now it is possible to calculate the size of the gun crossover, d gun. From Section 3.1, the distance between the gun crossover and the first condenser lens is 150 mm and the focal length of this lens is 1.4 mm. Then, thin lens theory gives

$$d = 0.285(150/1.41) = 30.5 \text{ u.}$$

This is 50% bigger than the manufacturer's claim of 20 u. However, the calculated number should be considered

accurate since the half angles involved are such that neither spherical aberration nor the Airy disk aberration makes a significant contribution to the imaging process.

2.3 Computer Calculations of the Theoretical Resolution

In chapter one, the following expression for the size of the probe was obtained (Equation 1.1).

$$d_s^2 = d_g^2 + \left(\frac{1.22\lambda}{\alpha_i}\right)^2 + \left(C_c \frac{\Delta V}{V} \alpha_i\right)^2 + \left(\frac{1}{2} C_s \alpha_i^3\right)^2 \quad .$$

The only two really important terms in the case of the thermionic gun are the demagnified Gaussian spot size d_g and the spherical aberration $1/2C_s\alpha^3$. The thin lens formula was assumed to apply to all the lenses in order to calculate the size and location of the image of the Gaussian spot d_g. The half angles were calculated in the same way. The data given in Section 2.1 was used as well as the experimental value for the gun crossover size determined in Section 2.2. The energy was taken to be equal to 100 kev because all the experiments were carried out at this energy. It was assumed that $\Delta v = 1$ ev. This value is the same as that used by Riemer (1972) and is approximately correct for normal filament temperatures but neglects the Boersch effect. A computer program was written using the expression above to calculate the theoretical resolution. The

values used for the second condenser lens strength were chosen so that the value of the theoretical half angles at the specimen covered the same range as the values measured in Section 2.2. Computer calculations were then carried out for condenser aperture sizes of 20 u., 35 u. and 55 u. The results of these calculations appear on graphs in chapter three where they are compared to the experimental results.

2.4 Methods of Measuring Resolution

There are a number of methods of measuring resolution. The most fundamental is to visually examine a test micrograph to find the closest spacing of two resolved points (eg. metal islands on a carbon film). Even then it is hard to be sure that the smallest spacing has been found by the visual scan.

Another possibility is to measure the width of a sharp edge. Here the edge can be regarded as a point object and the amount it is broadened as the resolution. The worse the resolution the more the edge will be broadened. This is the approach used by Gentch, Reimer et.al. (1973, 1974). In this procedure a picture is taken and either a visual check with an eyeglass or for a more accurate determination, a photodensitometer trace taken across the edge. The problem here is that there

are a number of transfer functions involved. These functions include the transformation of electrons to light, light to film exposure and film exposure to film density. In general these functions are non-linear. Because of this it is hard to get accurate results from this technique but rough results can be obtained.

A variation of this method was tried. It is possible to put the instrument in the line scan mode so that a trace whose vertical component is the transmitted intensity is displayed. A thin film of Mo was burnt forming MoO_3 crystals with very sharp edges. These crystals were deposited on a 1000 mesh grid. The instrument was then brought to focus and a picture taken so that the angle between the edge and the line scan would be known. It was then switched to line scan at fairly slow speed (at least 10 seconds per scan). It turned out that instead of a smooth monatomic transition across the edge, irregularities in the trace were observed.

It was concluded that this effect was caused by contamination. There are always organic molecules present in the microscope column due to the pump oil. Sometimes the specimen or external sources are also responsible. An interaction between these molecules and the electrons in the beam causes them to be ionized or cracked. They then migrate to the specimen

surface where a film of organic material is built up which tends to obscure specimen detail. The use of cold traps and cold fingers has made this problem unimportant in CTEM but in STEM, and particularily in this experiment due to repeated scans over the same area, it can be a dominating factor. It would have been possible to reduce the contamination by using shorter sweep times but then noise would have been a problem. Perhaps the line scan method would work if the vacuum was improved. However, without improvements to the instrument, the results obtained are meaningless.

Another attempt to measure resolution was made employing an optical diffractometer. A discussion of the diffractometer is given in Appendix I. A micrograph of a thin film with small metal (Pt Pd) islands on it was taken and put in the diffractometer. In principle this should have given rise to a circular transform whose size would be inversely proportional to the resolution in the micrograph. In practise this did not occur because the photographic emulsion used contained matting agent. The light would be scattered by the matting agent resulting in a circle of noise much larger than the transform. It was, however, possible to overcome this problem by printing onto a plate that had no matting agent.

However, even with no photographic plate in the diffractometer there was still a halo around the spot. It was presumed that this effect was due to optical defects in the diffractometer. Due to noise, the smallest details that it was reasonable to try to obtain in a micrograph were of a size such that the transform was a circle whose diameter was smaller than that of the halo. Thus it tended to be obscured by the halo. The number of details of this size tended to be small making the transform faint near its edge. In addition, it is impossible to take STEM micrographs without noise being present. This noise also has a transform which tended to further obscure the desired transform. In view of these factors this method was abandoned.

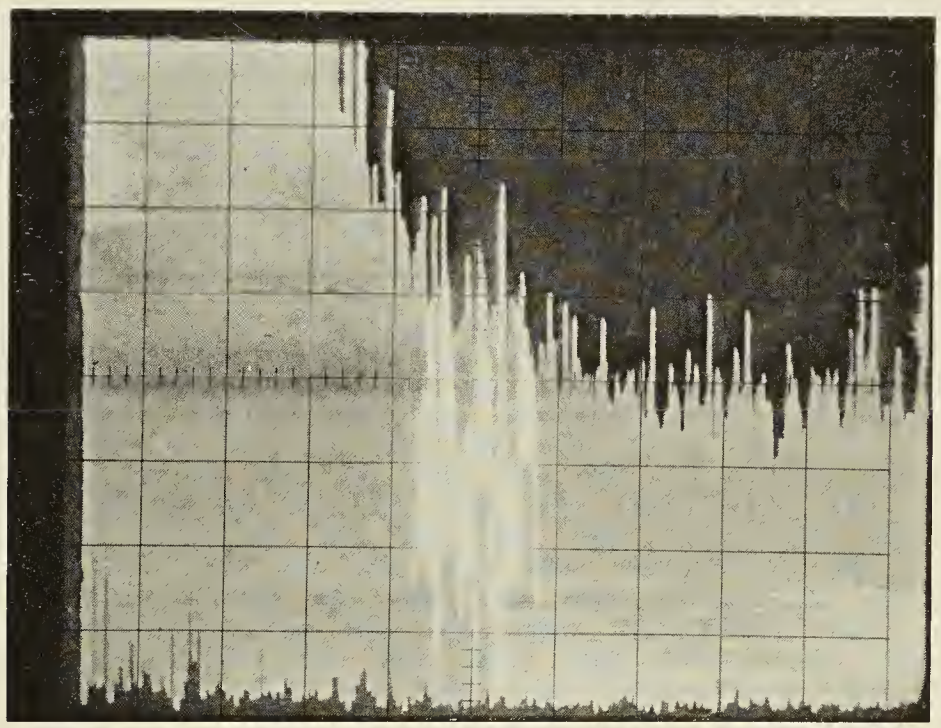
The way that the resolution was eventually measured was by spectrum analysis. A detail of a given size will give rise to a characteristic primary frequency in the electron detector which depends on the speed of the spot and size of the detail. If the output of the detector could be Fourier analysed then it would be possible to find the highest frequency present and thus get a measure of the resolution. In order to do this the output from the elctron detector was connected to a spectrum analyser. This is basically a band pass filter whose frequency varies with time. It was possible to set the width of the band to the desired value and this was

set as narrow as possible taking noise into account. The output from the analyzer was recorded on a storage oscilloscope. If a detail whose frequency was the same as the band pass filter was scanned its amplitude was recorded on the storage scope. It is necessary to note that the background noise of that frequency was also recorded on the oscilloscope.

The microscope was set to the desired operating condition and an objective aperture, of a size such that all undeflected electrons were transmitted, was inserted. The specimen with the Pt Pd islands on a thin film was put in and focussed on the STEM screen. Then the oscilloscope screen was cleared and the spectrum allowed to build up for one or two frames of one hundred seconds each. A polarioid picture was taken and the relevant data recorded. Plate II is an example of the pictures taken.

This example will now be analyzed. A signal generator with the frequency good to less than 1% error was used to analyze the pictures. The center frequency (frequency at the origin of the abscissa, see Plate II) was found to be 1300 Hz and the dispersion (one graduation on the abcissa) 215 Hz. Thus the frequency corresponding to the highest observed resolution was found to be 1635 Hz. However, about 10% error bar have to be taken into account because of noise and the width

Plate II. Typical spectrum.



of the band pass filter.

In order to convert this frequency into a resolution it is necessary to know the speed of the scanning spot. To find this one must know the time to scan one line, the magnification and the width of the picture. In order to find the time to scan one line the X output of the scanning box was connected to an oscilloscope and the time for one line found to be 1/15 of a second. The magnification was found to be 153,000 X by means of a magnification grating. The width of the viewing screen was found to be 100 mm. If r is the resolution, then

$$f = \frac{\text{speed of spot}}{\text{size of detail}} = \frac{\text{lines/sec} \times \text{width of screen/magnification}}{\text{r(resolution)}}$$

rearranging gives

$$r = \frac{\text{lines/sec} \times \text{width of screen}}{M \times f}$$

$$= \frac{15/\text{sec} \times 100 \text{ mm.}}{153{,}000 \times f}$$

$$= \frac{98{,}200}{f} \text{ Å} \quad .$$

Then in this case

$$r = 60 \text{ Å} \quad .$$

CHAPTER III

RESULTS AND CONCLUSIONS

In this chapter the results of the resolution as a function of incident half angle measurements are given. It was found that good agreement with theory is obtained when spherical aberration is not important but that agreement is poor where it is dominant. Some plausible explanations for this are considered. Also some preliminary work on resolution as a function of thickness is described and preliminary conclusions drawn. In addition a few general comments on problems encountered in the work are made. Finally there are some suggestions for further work.

3.1 Resolution as a Function of Incident Half Angle

i) Measurements

The resolution of the microscope was measured by the method outlined in chapter three. This was done for three different apertures (20 u., 35 u. and 55 u.) using nine different condenser lens strengths for each aperture. These lens strengths were converted into half angles using the measurements described in Section 2.2. Then the resolution was plotted as a function of half angle for a given aperture size. The computer predictions described in Section 2.3 were also plotted on the

graphs for a comparison. It was not possible to plot all the data on one graph since, although for a given second condenser lens setting the demagnified Gaussian spot size is always the same,the half angle at which it occurs is different due to different apertures. Therefore, three graphs were plotted which appear in Figures 3.1 to 3.3. The various contributions to the probe size are plotted as well as the total theoretical probe size. The points on the resulting curves are indicated by the following symbols:

⊙ demagnified Gaussian spot size contribution

\+ spherical aberration contribution

Δ diffraction limit contribution

× chromatic aberration size contribution

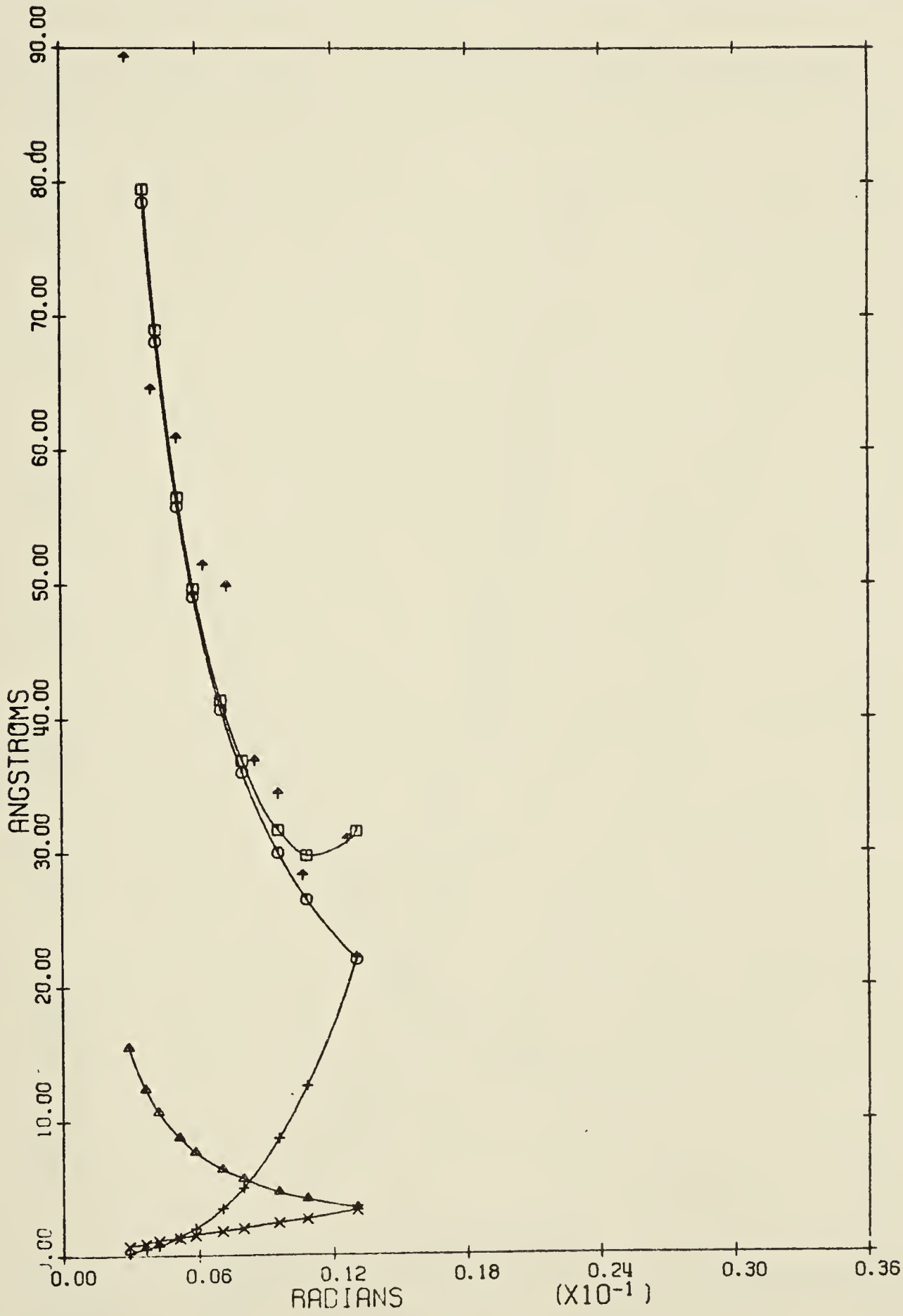
⊡ total theoretical probe size

↑ experimentally determined resolution

ii) Comments on the Measurements of Resolution as a Function of Incident Half Angle

The results of the resolution as a function of angle measurements are interesting. In all three graphs, at low angles, where it may be expected that the size of the spot will be dominated by the demagnified Gaussian spot size, the agreement with theory is reasonable. (The error bars on the experimental points are about 10%

Figure 3.1. Resolution with 20 u aperture and C_s= 2.0 mm.

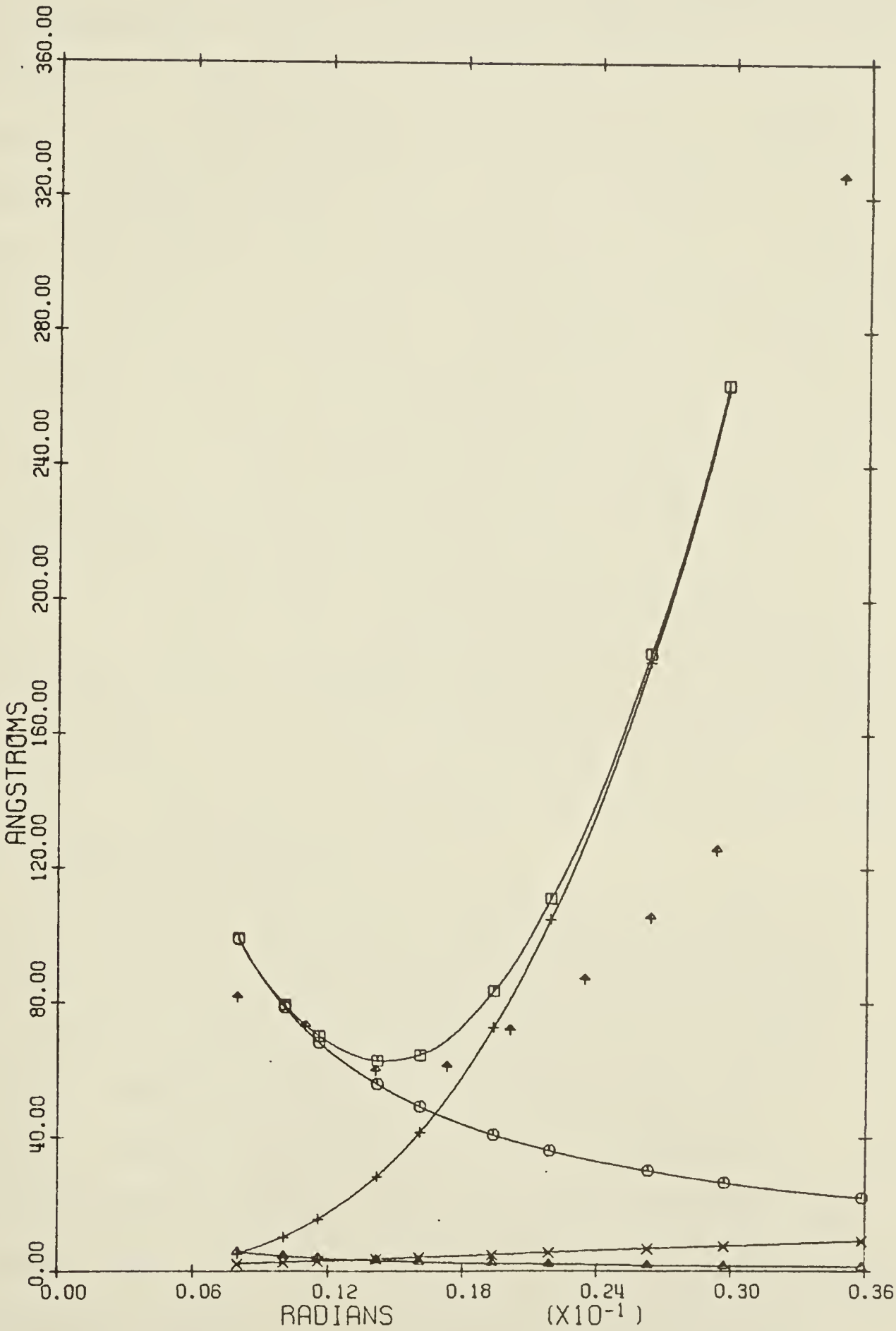


RESOLUTION VS. SPECIMEN ANGLE

Figure 3.2. Resolution with 35 u aperture and C_s = 2.0 mm.

RESOLUTION VS. SPECIMEN ANGLE

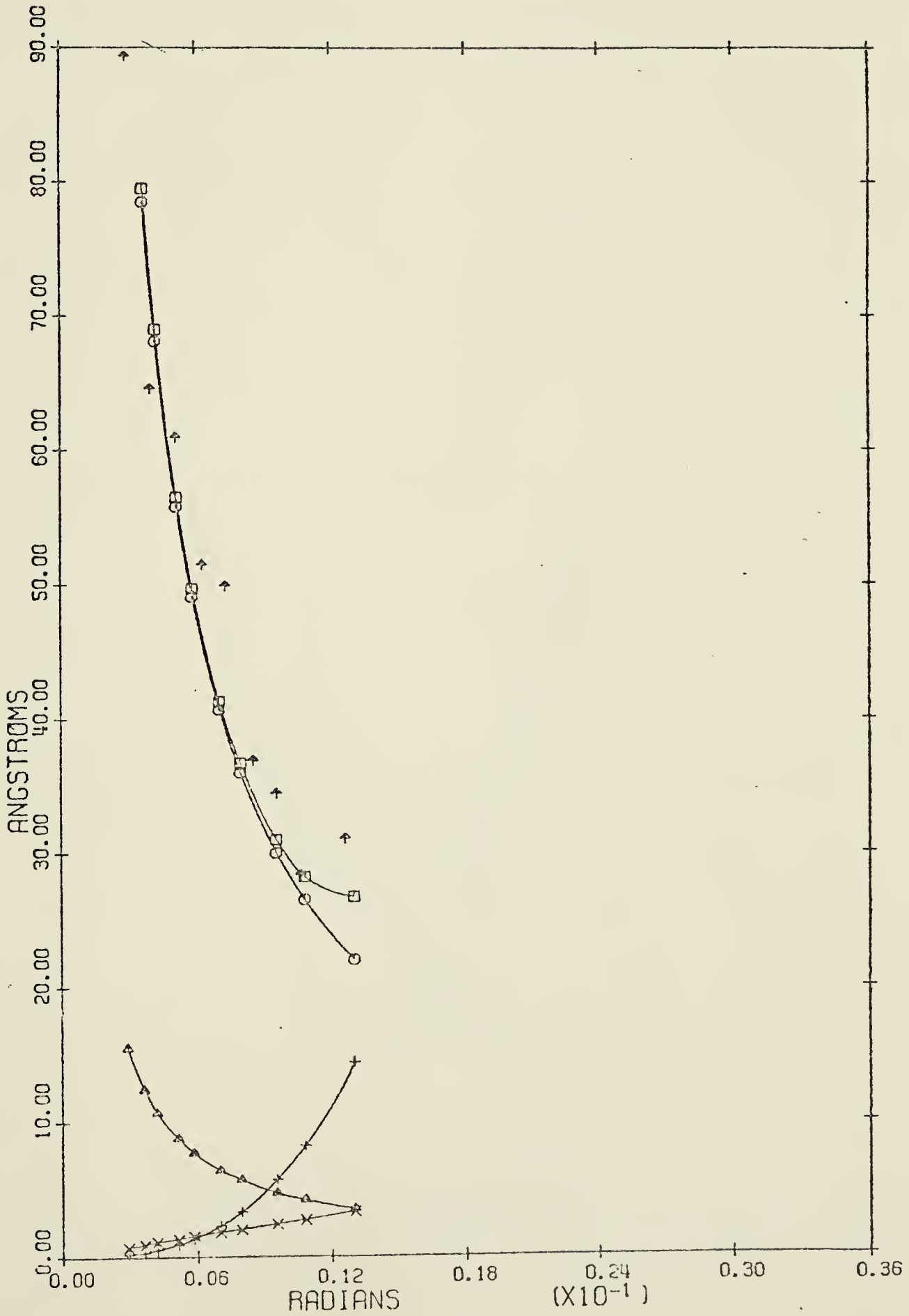
Figure 3.3. Resolution with 55 u aperture and C_s = 2.0 mm.



due to uncertainty in determining the frequency corresponding to the best resolution.) This is gratifying as the manufacturer claims the Gaussian spot size at 100 kv to be 20 u. as opposed to the measured 30 u. However, as can be seen from Figures 3.2 and 3.3, the agreement at large angles is poor with the 35 u. and 55 u. apertures. This means that the effect of spherical aberration is not as great as expected. In addition, in Figures 3.2 and 3.3, the minima of the curves of the measured points tend to fall at higher angles than the predicted locations also indicating that the effect of spherical aberration is less than what the theory suggests. In order to investigate this possibility, the theoretical curves have been recalculated using C_s = 1.3 mm instead of C_s = 2.0 mm. The results of this are shown in Figures 3.4 to 3.6. The fit is much better and the minima move over to agree with experiment.

Let us consider some reasons why the effect of spherical aberration may be less than the theory predicts. One possible reason might be that the effect is due to a smaller effective half angle than was measured. However it is difficult to think of any plausible reason why this would be the case since the convergent beam technique gives the exact conditions at the specimen without any complicated assumptions or

Figure 3.4. Resolution with 20 u aperture and C_s = 1.3 mm.

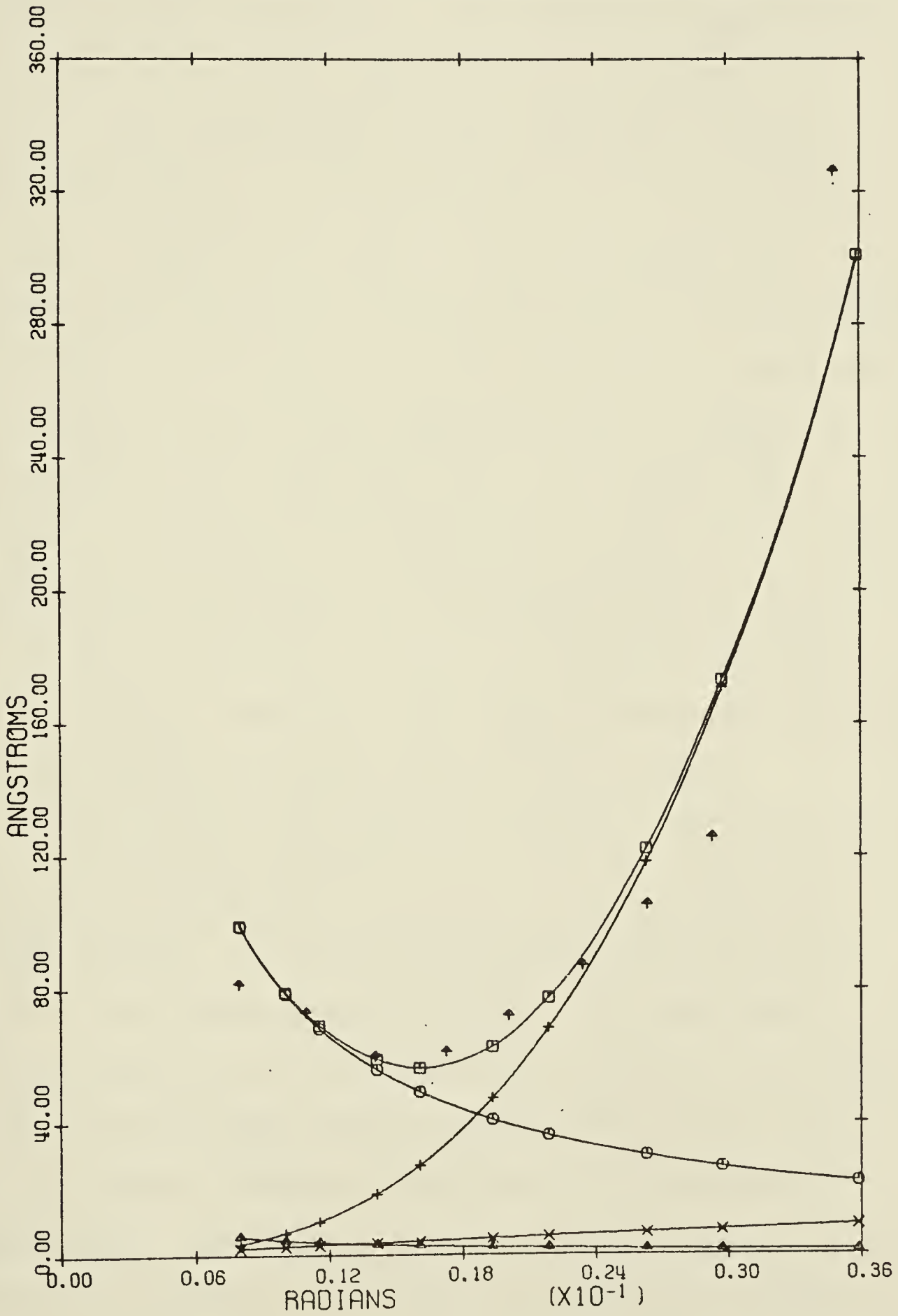


RESOLUTION VS. SPECIMEN ANGLE

Figure 3.5. Resolution with 35 u aperture and C_s = 1.3 mm.

RESOLUTION VS. SPECIMEN ANGLE

Figure 3.6. Resolution with 55 u aperture and C_s = 1.3 mm.



RESOLUTION VS. SPECIMEN ANGLE

calculations. It should be noted that the expression for spherical aberration is actually the diameter of the disk of least confusion. It is shown in Appendix III that the size of the disk of least confusion is $1/2\ C_s\alpha^3$. One possibility is that the value of the coefficient $1/2\ C_s$ in the expression $d_s = 1/2\ C_s\alpha^3$ is too large. However, nowhere in the literature are values of the lead coefficient less than 1/2 to be found for the size of the disk of least confusion as compared to the radial error in the Gaussian plane. Another possibility is that the value of C_s is smaller than that used. Comparison with tabulated values of the lens parameters reveals that it is unlikely that C_s is less than 2.0 mm if the focal length is 3.5 mm as the manufacturer states. However, since this is not experimental data the possibility exists that it is in error.

This leaves differences in the way the disks are defined and the method of addition of error disks as possible causes. The demagnified Gaussian spot d_g is defined as the width at half height. The spherical aberration disk d_s is defined as the diameter into which all the electrons that leave a point source will fall. These two disks, defined in different ways, are added in quadrature. The effect is that the spot size is essentially defined as the width at half height where the Gaussian spot size is dominant and spherical

aberration is unimportant. However, when spherical aberration is dominant and the Gaussian spot size unimportant the resolution is essentially defined as a circle into which all the electrons fall. In between these two extremes a somewhat intermediate result is obtained. However, the criterion that determines the resolution of the probe is the width at half height, so it would be expected that this method tends to over-estimate the effect of spherical aberration since the full diameter is used instead of the width at half height. The experimental results seem to bear this out. In order to do the theoretical calculation properly, it would be necessary to trace a large number of rays and take the resolution as the width at half height.

3.2 The Resolution as a Function of Thickness

i) The Initial Attempt

The initial attempt to measure the relationship between thickness and resolution was made with Bismuth specimens. The specimens were made by depositing Bismuth on a glass slide to form a thin film while the thickness was measured by simultaneously depositing Bismuth on a quartz crystal and observing the changes in natural frequency due to the increased mass. The specimen was placed in the microscope and the details brought to the

sharpest possible focus. A micrograph was taken which was then visually scanned to determine what the closest spacing of any two resolved details was. This was taken to be the resolution.

Because of the top-bottom effect, the best resolution will always be at the top of the specimen. Due to this, the method used will result in every micrograph being taken with the focus at the top of the specimen and thus there was no change in resolution regardless of the thickness of the specimen. Because of this a new approach was necessary.

ii) The Method of Measurement

The specimen finally used in the resolution as a function of thickness measurements consisted of a thin film of silicon with germanium islands deposited on one side. The silicon disk had a hole near the middle and the silicon formed a thin wedge in the vicinity of the hole (this will be shown later). Direct measurements of the thickness could not be made but the distance from the edge of the wedge was measured by reading the micrometer on the specimen translate drive. From this it was possible, assuming a uniform wedge angle to obtain the relative thicknesses. At selected distances into the wedge, STEM micrographs were taken in order to determine the resolution. In addition, at every

position that resolution was measured, a diffraction pattern was taken. In order to get some measure of the thickness, the average current reaching the detector was measured using a Faraday cage and electrometer. For comparative purposes high resolution CTEM pictures were taken at each position. Low magnification CTEM pictures were also taken near the edge in an effort to determine the wedge angle. The wedge was then turned over and the experiment repeated.

It should have been possible from the low magnification CTEM pictures to determine the wedge angle at its edge from the extinction contours. If the wedge was uniform, these contours would be parallel to the edge and spaced evenly. However, neither of the above conditions was fulfilled indicating that the wedge did not increase uniformly in thickness with distance from the edge for small thicknesses.

The above result shows that the wedge did not have an absolutely linear relationship between distance into the wedge and thickness for small thicknesses. In spite of this it was possible to show that a linear relationship existed throughout the wedge although it was likely that small deviations were present throughout. It was thus possible to obtain the relative thicknesses of the places where the resolution micrographs were taken.

This was done by means of the current measurements. It was assumed that due to scattering the current would fall off exponentially with thickness. (There are problems with this assumption due to dynamical scattering. However, most of the measurements were taken far past the point where the extinction contours had disappeared. Under these conditions, particularly for a two beam situation, this assumption is not totally unreasonable.) Therefore, if the wedge was linear, there would be a linear relationship between distance and the logarithm of the current. The data was plotted to check this and the results appear in Figure 3.7. It would appear that the wedge was close to being uniform on the larger scale although one line becomes non-linear in the thick part of the wedge.

The micrographs were examined with a Polaron 10 x magnifier with scale and the spreading of the edges of the islands taken to be the resolution in both the CTEM and STEM modes. An attempt was made to plot all the data on one graph by assuming that equal current meant equal thicknesses. However, it was found that the two CTEM curves did not coincide and that the diffraction effects were different at what was supposed to be the same thickness. It was concluded that this was due to the orientation dependence of the scattering factors and that possibly the orientation was not exactly the same

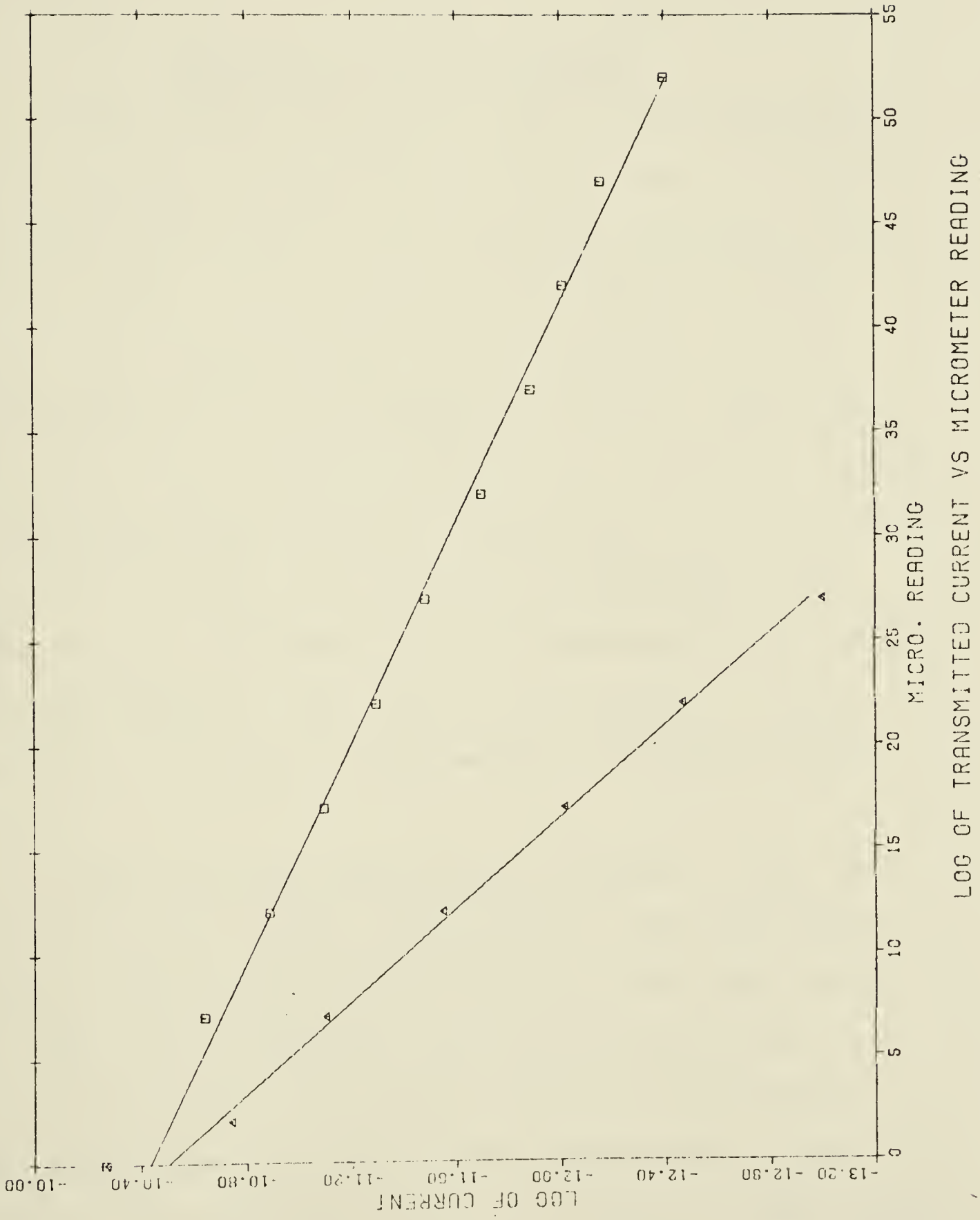


Figure 3.7. Linearity of the wedge.

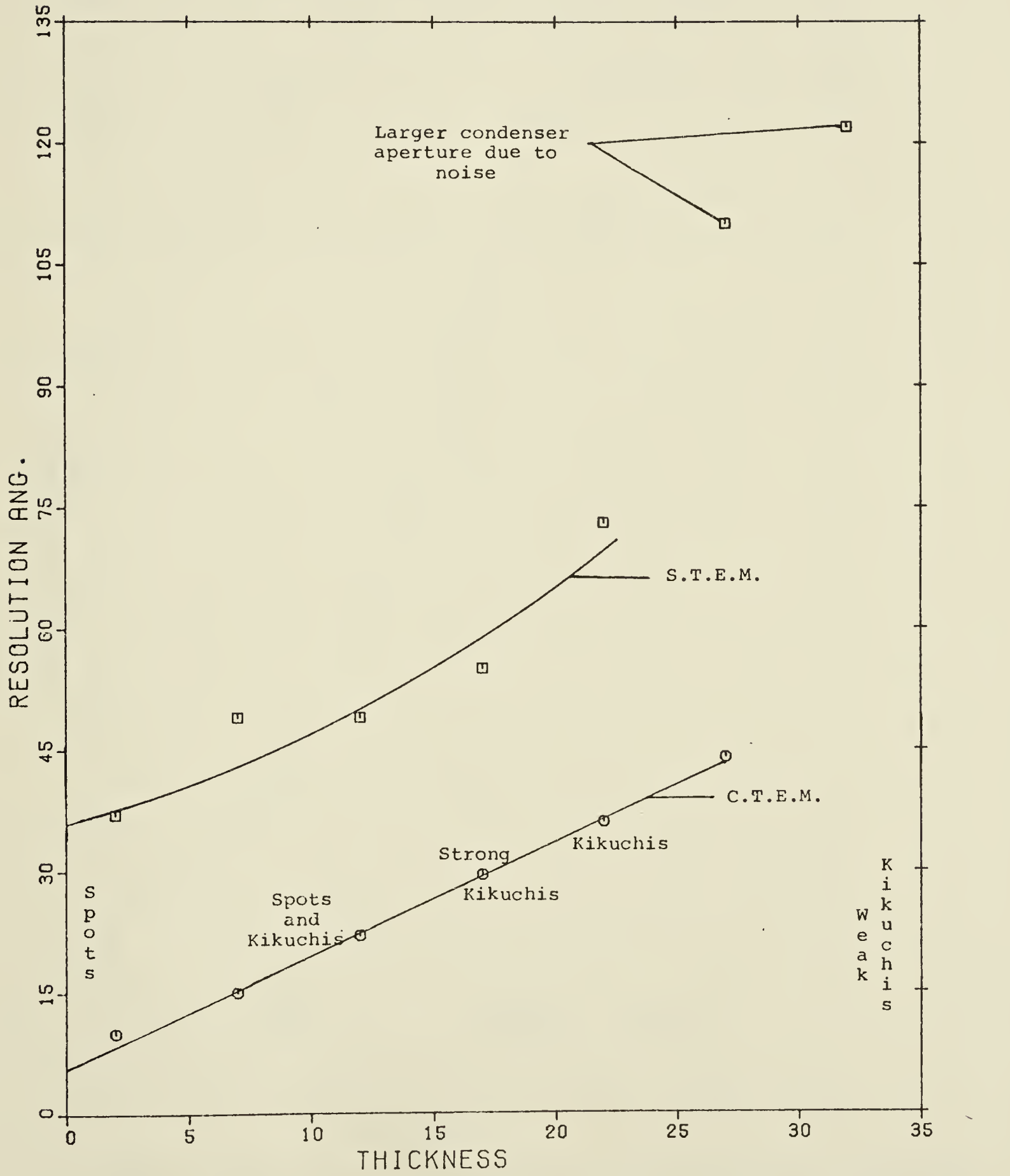
in both cases. Therefore, the results are presented in Figures 3.8 and 3.9 with no attempt to provide a relative scale.

iii) Comments on the Measurements

The diffraction patterns can give an indication of the amount of scattering in the specimen. If there are only spots in the pattern then the specimen is thin enough so that little inelastic scattering has taken place. If Kickuchi lines are also present, inelastic scattering has begun to disperse the beam. When there are only strong Kickuchi lines present, the beam has been badly dispersed by inelastic scattering. As the crystal gets thicker still, the variation of intensity with respect to the incident direction decreases to the point that the Kickuchi lines weaken and disappear. These different regions have been marked on Figures 3.8 and 3.9.

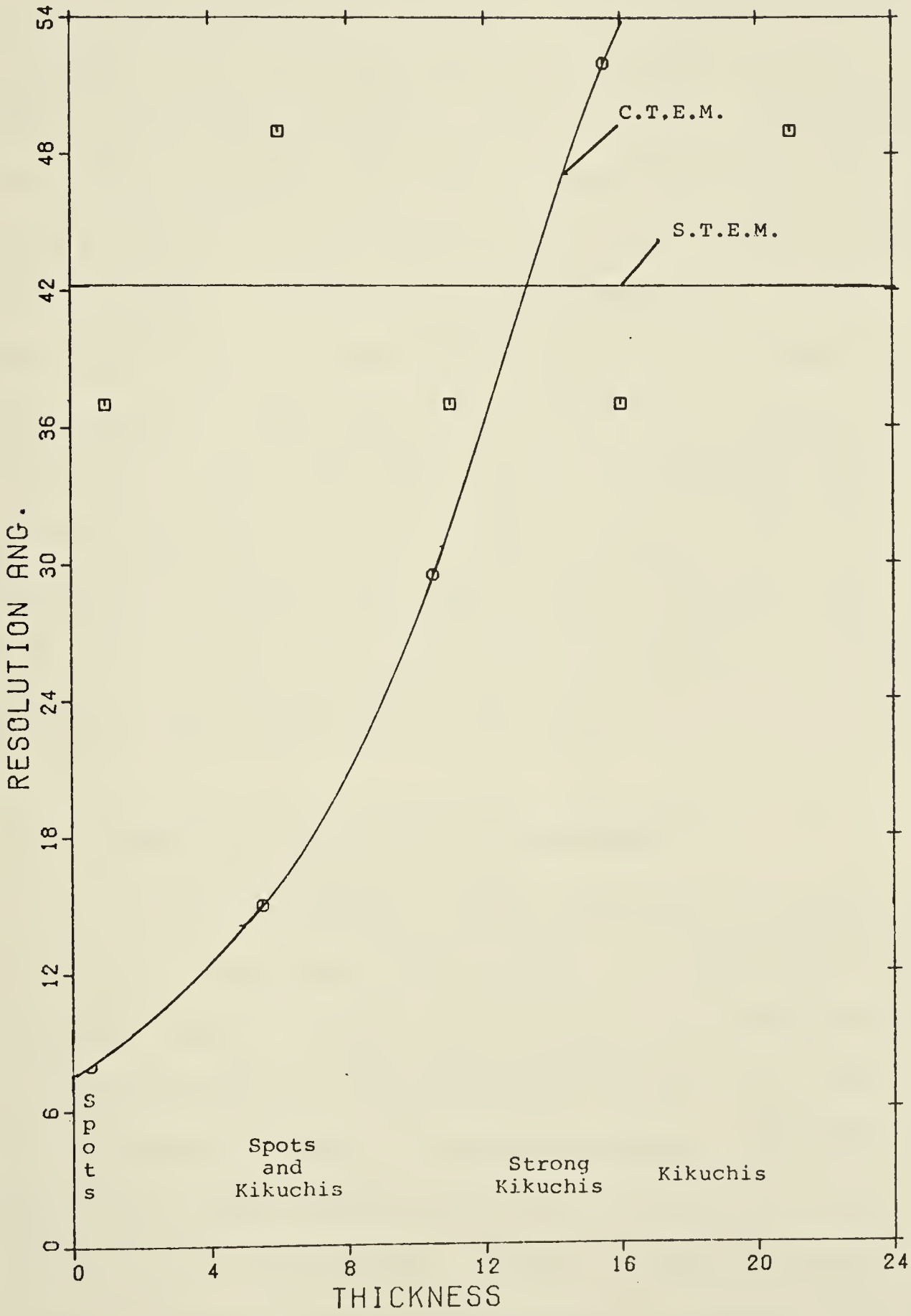
There is some question as to whether or not the resolution measurements are reliable. The same objections to the methods of Gentch et.al. also apply here i.e. the various transfer functions are non-linear. There are problems controlling exposure and development conditions. In addition the observation of the widths of the Germanium island is subject to large errors due to observer judgement and the small widths being

Figure 3.8. Thick specimen resolution with details on the bottom.



RESOLUTION AS A FUNCTION OF THICKNESS IN SILICON

Figure 3.9. Thick specimen resolution with details on the top.



RESOLUTION AS A FUNCTION OF THICKNESS IN SILICON

measured. However in spite of this it is still possible to draw some preliminary conclusions.

The top-bottom effect is clearly in evidence. In Figure 3.8 the resolution is dependent on the thickness in STEM but in Figure 3.9 it is not. It is obvious that CTEM is better than thermionic STEM with the islands on the bottom of the specimen but due to noise problem and the rough nature of the measurements it is not possible to say whether the CTEM and STEM curves are converging or diverging. The thing that limited the thickness of specimen that could be observed in the STEM mode was always noise. In fact, some of the STEM points in Figure 3.8 had to be taken using a larger probe size than the minimum that was possible in order to obtain sufficient electrons at the detector to take a micrograph.

3.3 Comments on Problems Encountered

It was found that rather than contaminating uniformly, the specimen would appear to contaminate in vertical bands. It was found that this problem was due to the presence of stray magnetic fields which were due to fluorescent lights in adjacent rooms. Probably this was due to a pick up by the system electronics that controlled the scan coils. This in turn led to the probe not moving in a uniform fashion. Thus different parts of the micrographs were at different magnifications.

Since the probe would spend more time in some areas than others, these areas would contaminate more quickly resulting in the observed bands. This problem was overcome by working at those times when these lights could be turned off.

Contamination was another problem that plagued the project. Although the decontamination system was quite adequate for CTEM it failed utterly in STEM. It has been shown (Warren A. Knox, 1974) that under scanning conditions, contamination builds up quite rapidly. The rate of contamination in the experiments was specimen dependent suggesting that a major source of the organic molecules could be the support film for the specimen. In addition, it was noticed that sometimes with a specimen supported by an organically based support film that an area that had been scanned would appear more transparent after one pass than its surrounding suggesting that some material had been given off. After this though, it would contaminate rapidly.

3.4 Further Work

One important project would be the development of more effective anticontamination measures for use in STEM. At the same time, the effect of the specimen on its own contamination should also be investigated.

Some work should be done to investigate the reason for the smaller effective coefficient of spherical aberration. This project should include repeating the present experiments in order to get better statistics on the value of the resolution and finding a way to measure C_s in the scanning position. In addition, a theoretical study of the type suggested at the end of Section 3.2 ii) should be undertaken.

Subsequent to the completion of the experiments on resolution as a function of thickness presented in this thesis, additional results have appeared in the literature. Fraser and Jones (1975) have reported no increase in penetration in gold using thermionic STEM. Fraser et.al. (1977) using silicon have concluded that there is no advantage to STEM even with a field emission gun for all crystalline specimens. They also concluded that thermionic STEM is quite inferior to CTEM. However, Easterling (1976) reports that increased penetration was obtained using thermionic STEM and specimens of tungsten, stainless steel and aluminum. It would appear that due to this contradiction in the literature further resolution as a function of thickness experiments are justified particularly on low atomic weight, crystalline materials.

3.5 Conclusions

It has been found that in thin specimens the theoretical resolution does not agree with the experimentally

determined resolution when this is primarily determined by spherical aberration - resolution is better than predicted at high angles. A possible explanation has been found based on the fact that the disk of confusion is not measured as width at half height. For thick specimens some experiments were done that enable a few preliminary conclusions to be drawn. The top-bottom effect is clearly evident and it is clear that for details on the bottom of the specimen, thermionic STEM is inferior to CTEM. However, there is a great deal of work that remains to be done to confirm this.

BIBLIOGRAPHY

Cowley, J.M., Image Contrast in a Transmission Scanning Electron Microscope, Applied Physics Letters 15, 58 (1969).

Cowley, J.M., Diffraction Physics, pp. 51-55, North-Holland Publishing Co., Amsterdam (1975).

Crewe, A.V., Scanning Electron Microscopes: Is High Resolution Possible? Science 154, 729-738 (1966).

Crewe, A.V., Wall, J. and Welter, L.M., A High Resolution Scanning Electron Microscope, Journal of Applied Physics 39, 5861-5868 (1968).

Crewe, A.V. and Wall, J., Contrast in a High Resolution Scanning Transmission Electron Microscope, Optik 30, 461-474 (1970).

Cucack, N., The Electrical and Magnetic Properties of Matter, pp. 17-20, 31-32, 42-43, 55-56, 60-78, Longman, Green and Co.,Ltd., London (1967).

Easterling, K.E., International Metal Review (in press) (1976).

El-Kareh, A.B. and El-Kareh, J.C.I., Electron Beams, Lenses and Optics, Volumes 1 and 2, Academic Press, New York (1970).

Fraser, H.L. and Jones, I.P., A Note on the Increase in Usable Foil Thickness in STEM, Philosophical Magazine 31, 225 (1975).

Fraser, H.L., Jones, I.P. and Loratto, M.H., Limiting Factors in Specimen Thickness in Conventional and Scanning Transmission Electron Microscopy, Philosophical Magazine 35, 159-175 (1977).

Von Gentsch,P. and Reimer, L., Messungen zum Auflösungsvermögen in der Raster-Transmissionselektromenmikroskopie dicker Objeckte, Optik 37, 451-454 (1973).

Gentsch,P., Hilde, H. and Reimer, L., Measurement of the Top-Bottom Effect in Scanning Transmission Electron Microscopy of Thick Amorphous Specimens, Journal of Microscopy 100, 81-92 (1974).

Haine, M.E., The Electron Microscope - The Present State of the Art, pp. 115-141, E. & F.N. Spon Limited, London (1961).

Hawkes, P.W., Electron Optics and Electron Microscopy, pp. 45-47, 87-88, 124-150, 151-152, Taylor and Frances Ltd., London (1972).

Howie, A., The Reciprocity Theorem in Electron Microscopy and Diffraction, Proceedings of the Fifth European Congress on Electron Microscopy, pp. 409-413, University of Manchester (1972), published by the Institutes of Physics, London.

Knox, W.A., A Study of Contamination caused by a Very Narrow Electron Beam, 32 Annual Proceedings of the Electron Microscopy Society of America,

Saint Louis, Missouri (1974), L.J. Arceneaux (ed.).

Koike, H., Ueno, K. and Watanbe, M., Scanning Devices Combined with Conventional Electron Microscope, Septiéme Congrés International de Microscopie Électronique, Grenoble (1970), Pierre Favard (ed.).

Langmuir, D.B., Theoretical Limitations of Cathode Ray Tubes, Proceedings of the Institute of Radio Engineers, pp. 977-991 (1937).

Oatley, C.W., The Scanning Electron Microscope - Part I The Instrument, Cambridge University Press, Cambridge (1972).

Ong, P.S., A High Resolution Electron Microscope for Conventional Imaging and Scanning Mode of Operation, Fifth International Congress on X-Ray Optics and Microanalysis, Tübingen (1968), V. Möllenstedt and K.H. Gaukler (eds.).

Reimer, L., Physical Limits in Transmission Scanning Electron Microscopy of Thick Specimens, Scanning Electron Microscopy/1972, Proceedings of the Fifth Annual Scanning Electron Microscope Symposia, Om Johari and Irene Corvin (eds.).

Takashima, S., Hashimoto, H. and Kimoto, S., Resolution of a Transmitted Electron Image formed by a Scanning Transmission Microscope, Septiéme Congrés International de Microscopie Électronique, Grenoble, (1970), Pierre Favard (ed.).

Thorton, P.R., Scanning Electron Microscopy, pp. 8-9, Chapman and Hall Ltd., London (1968).

Valentine, R.C., The Response of Photographic Emulsion to Electrons, Advances in Optical and Electron Microscopy, Vol. 1, pp. 187, R. Buren and V.E. Cosslett (eds.).

Appendix I

The Diffractometer

In the course of the experiments, a diffractometer was built. It is a well known result that a circular object of diameter D will give rise to a diffraction pattern whose central disk (called the Airy disk) is of angular radius θ where

$$\theta = 1.22\ \lambda/D \quad .$$

Since the rings in the pattern are much weaker than the central disk, it is reasonable to expect that if a specimen with many different sized circles were put into a light beam that the rings would not contribute significantly to the resulting pattern and that it would be possible to observe a disk corresponding to the smallest circles present. It was to take advantage of this that the diffractometer was built.

Figure A.1 is a diagram of the diffractometer. A four meter triangular optical bench fastened to a piece of plywood carried most of the optical components of the system. The illumination was provided by a ruby red laser of low intensity. This was mounted on a platform attached to the optical bench. This platform permitted the laser to be tilted in the vertical plane of the optical bench and rotated in the horizontal

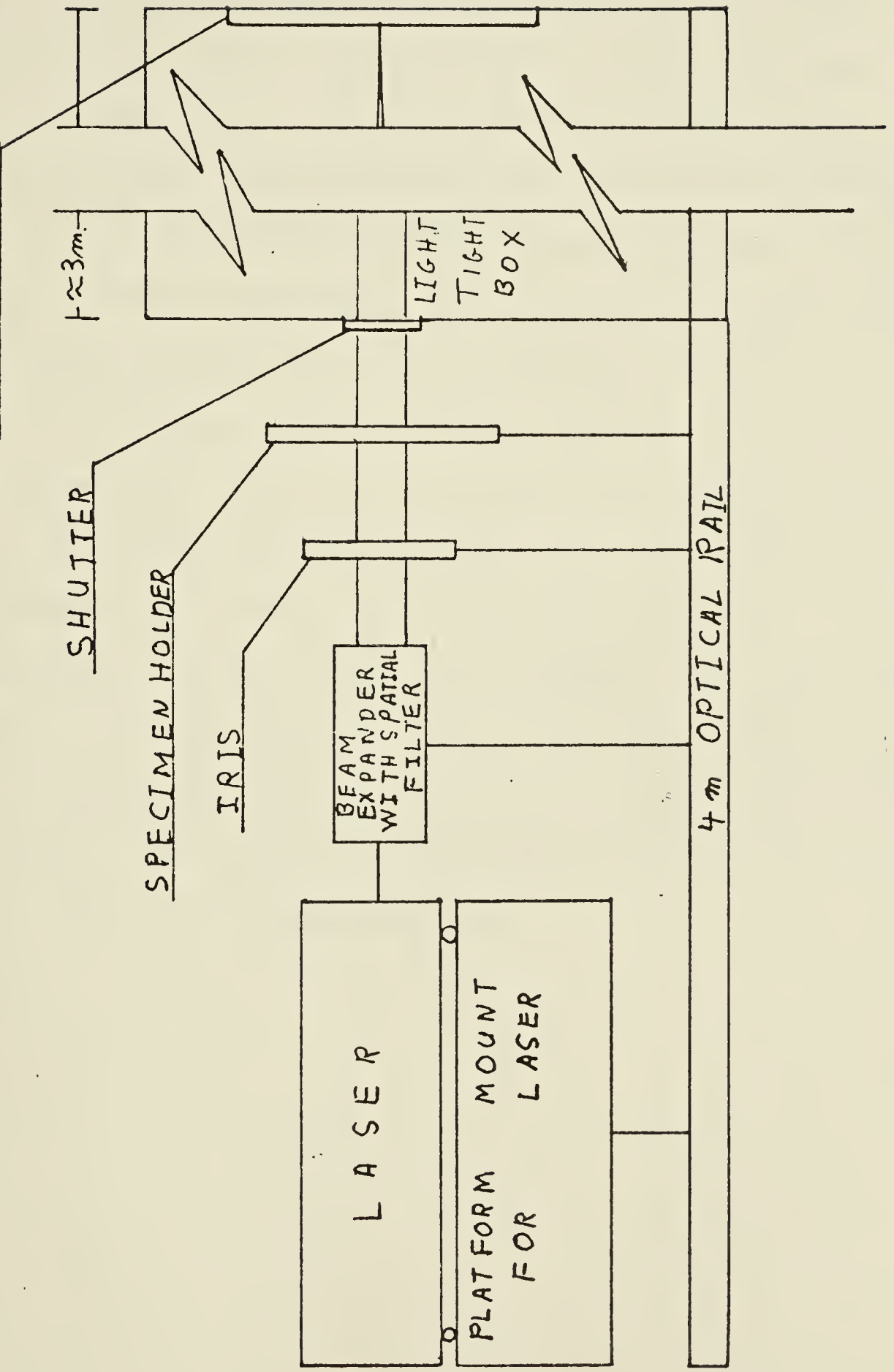


Figure A.1. Diagram of the diffractometer.

plane for alignment purposes. The illumination then went through the beam expander with spatial filter. The spatial filter was necessary in order to clean up the beam due to the presence of off axis rays in the illumination. This component was mounted on vertical and horizontal translate devices for purposes of alignment. Rather than being set for parallel illumination the beam expander was set to focus on the recording surface (ground glass or film). The next component was an adjustable iris. Following was the specimen holder that was adjustable vertically and horizontally in order to select any desired area for diffractometry. Finally there was a 'light tight' box slightly over three meters long. This was painted flat black inside and out. At one end was a shutter and at the other end a recording surface. This consisted of either a ground glass screen or a film for taking pictures.

Appendix II

Alignment

One of the most important aspects of microscope operation when measuring resolution is the microscope alignment. Improper alignment will result in degraded resolution so it is important that the method used give the best possible alignment. Several methods were used but the one eventually adopted is described below.

a) Centering the gun. The first condenser lens was set to its strongest excitation (number four position) and the spot centered with the align translate control. The first condenser lens was then set at its weakest excitation (number one position) and the gun translate used to center the spot. This cycle was repeated until the spot did not move. The explanation of this procedure is as follows. If the gun crossover is off center the first condenser crossover is also off center. It can be made to appear to be on center by a suitable adjustment of the align translate control. When the focal length of the first condenser lens is increased, the crossover will move further off center resulting in the spot moving on the screen. Adjusting the gun translate to bring it back will move the gun towards the axis until the crossover is the same distance from the axis as it was when the lens was strongly excited. By

cycling, the method converges on the proper alignment.

b) Setting the gun tilt. The filament current was reduced until the filament 'image' was obtained (a spot with a halo around it). This pattern was then made symmetrical by changing the gun tilt so that the halo was uniform around the spot. The explanation of this is that when the filament current is reduced the emission current is also reduced since the filament is cooler. This in turn causes the bias to decrease because auto biasing is used. This allows the electric field to penetrate further into the Wehnelt cylinder. If the tilt of the Wehnelt is not parallel to the axis the electric field will penetrate unevenly resulting in deeper penetration on one side of the filament to the other. This results in an asymmetric halo. It is important to have the correct tilt in order to obtain maximum brightness.

c) Correcting the condenser astigmatism. The condenser stigmator amplitude was turned to zero. The spot was focussed down to its smallest size on the screen and the focus changed slightly so it assumed its maximum asymmetrical shape. Then the stigmator was used to produce an opposite astigmatism and the amplitude reduced till the spot became a circle.

d) Setting the bright field tilt. The microscope was focussed on a test specimen. The high voltage

wobbler, which changes the accelerating voltage periodically, was turned on and the tilt adjusted until the image no longer moved. The explanation for this is that if the tilt is not parallel to the axis the cone of illumination will move on the specimen. Since the focal length of the lenses is energy dependent, the image will also move. This step is necessary in order to be able to do step e).

e) Centering the intermediate lenses. The microscope was set to diffraction and the illumination spread to form a caustic. This was brought to the center of the screen with the intermediate lens translate controls. The caustic is caused by aberrations in the objective lens. This step was necessary in order to do f).

f) Setting the scanning tilt. The second condenser lens control was set at four positions from its maximum and the microscope set to dark field. The objective lens current was increased until a caustic appeared with the magnification set to 5000X. This was then brought to the center with the dark field tilt. This method of setting the tilt is not very accurate but since the tilt coils are approximately in the front focal plane of the lens under scanning conditions, the tilt does not need to be particularly accurate.

g) The next step is to increase the objective current to the scanning value of approximately 500 mA.

h) Centering the intermediate lens. This procedure is the same as in e). This is necessary because the objective lens does not behave the same at full strength as it does at the CTEM strength used before. This prepares the way for step i).

i) Insert the objective aperture. This was then moved so that it was in the center of the beam.

If the scanner was now turned on, the projector lens off, the specimen inserted, the desired objective aperture put in and the screen lifted, the microscope was ready for operation.

Appendix III

Guns and Lenses

In chapter one it was shown that in thermionic STEM the brightness of the gun and the spherical aberration coefficient restrict the available resolution. This appendix fills in some details that were omitted and in addition contains some other material about the principles and operation of electron guns and magnetic lenses.

In Section 1, the characteristics of thermionic guns are considered with a special emphasis on the maximum brightness attainable. The brightness relationship is derived in two parts. In the first part, an expression is obtained for the current density available from a hot cathode. This derivation is based on material from Cusak (1967). The second part starts with the expression for the current density and develops a relationship which gives the maximum possible brightness. This part follows the original proof due to Langmuir (1937). Finally, some practical aspects of electron gun operation are considered with a special emphasis on gun biasing.

Section 2 is devoted to the lens characteristics. The entire section contains standard electron

optial material and is based on El-Kareh and El-Kareh (1970). First the equation of motion for an electron in an axial symmetric field is derived for a special case. From this equation it is then shown how spherical and chromatic aberration come about.

Section I: The Gun

The electron gun is one of the most important parts of any STEM instrument. As we shall see, thermodynamic considerations place an upper limit on the brightness of any thermionic gun. This in turn imposes a practical limit on the size of the aperture that one can use because of noise limitations. It is important that the gun be stable against fluctuations in operating voltage or chromatic aberration will degrade the resolution. It is also necessary to stabilize the gun so that the emitted current due to changes in filament temperature does not result in artifacts in the image.

First let us consider a line of reasoning that shows the existence of an upper limit on the brightness of a thermionic gun. Let us consider an infinite half plane of metal at temperature T with work function ϕ and Fermi level E_F. Let us assume that the electrons in metal form a free electron gas. Using these assumptions, the Schrodinger equation and imposing periodic boundary conditions on the solutions, it can be shown

that the volume per state in phase space is h^3. (For a discussion of this point see Cusack, 1967). Then the number of states between p and p+dp is $2dp_x\ dp_y\ dp_2/h^3$ (where the factor 2 takes care of the two possible electron spins). Let n(p) be the number of electrons between p and p+dp. Then

$$n(p)\ dp_x\ dp_y\ dp_z = 2\ \frac{f(E)}{h^3}\ dp_x\ dp_y\ dp_z$$

where f(E) is the Fermi-Dirac distribution function. There is a small potential hump at the surface; however, in order to simplify this discussion, it will be assumed that every electron arriving at the surface with a momentum perpendicular to the surface great enough to overcome the work function will escape. Then the number of electrons n that escape per unit time is

$$n = \int_0^\infty \frac{p_x'}{m}\, n(p_x')\ dp_x' \quad , \quad p_x' = [p_x^2 - m(E_F + \Phi)]$$

where p_x is the momentum perpendicular to the surface after escape and p_x' the momentum in the same direction before escape. Notice that

$$n(p_x') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} n(\vec{p}')\ dp_y\ dp_z$$

$$= \frac{2}{h^3}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \frac{dp_y\ dp_z}{\exp([(p_x'^2+p_y^2+p_z^2)/2m + \Phi]/k_BT) + 1} \quad .$$

If we do some calculations and let the current density at the surface be i we find that

$$i = en = \frac{4n\, m_e k_B T}{h^3} \int_0^\infty \log(1 + \exp[-(\varepsilon_x+\Phi)/k_B T])\, d\varepsilon_x \ .$$

Now $\exp[-(\varepsilon_x+\Phi)/k_B T]$ is always much smaller than unity since in any real case that may be encountered in a thermionic electron $\Phi/k_B T >> 1$. Therefore we can carry out a series expansion of the log term inside the integral and keep only the first order terms. If we do this and integrate we obtain

$$i = \frac{4n\, m_e k_B^2 T^2}{h^3} e^{-\Phi/k_B T} \qquad \text{(A.1)}$$

for the emitted current density from a flat surface.

The rest of the proof follows the same lines as the original proof due to Langmuir. The Fermi-Dirac distribution function which governs electrons in the hot cathode is

$$f(E) = \frac{1}{\exp[(\varepsilon-\varepsilon_F)/k_B T] + 1}$$

where ε is the energy of the individual electron. For those electrons that escape $\varepsilon > \varepsilon_F+\Phi$ and this makes $\exp[(\varepsilon-\varepsilon_F)/k_B T] >> 1$ for all operating temperatures. Therefore the energy distribution of the escaped electrons is

$$f(\varepsilon) \simeq e^{-\varepsilon'/k_B T}$$

where $\varepsilon' = \varepsilon - \varepsilon_F + \Phi$. This is the same as the Maxwell-Boltzmann distribution function. Notice that it is isotropic. From this function it can be shown from a general statistical argument that the number of electrons per unit volume with energy between ε and $\varepsilon+d\varepsilon$ (let us call this $n(\varepsilon)d\varepsilon$) is

$$n(\varepsilon)d(\varepsilon) = 2\,\frac{N}{V}\,\frac{\varepsilon^{\frac{1}{2}} e^{-\varepsilon/k_B T}}{\sqrt{n}\ (k_B T)^{3/2}} \quad .$$

Then the current of electrons di with energy between ε and $\varepsilon+d\varepsilon$ is

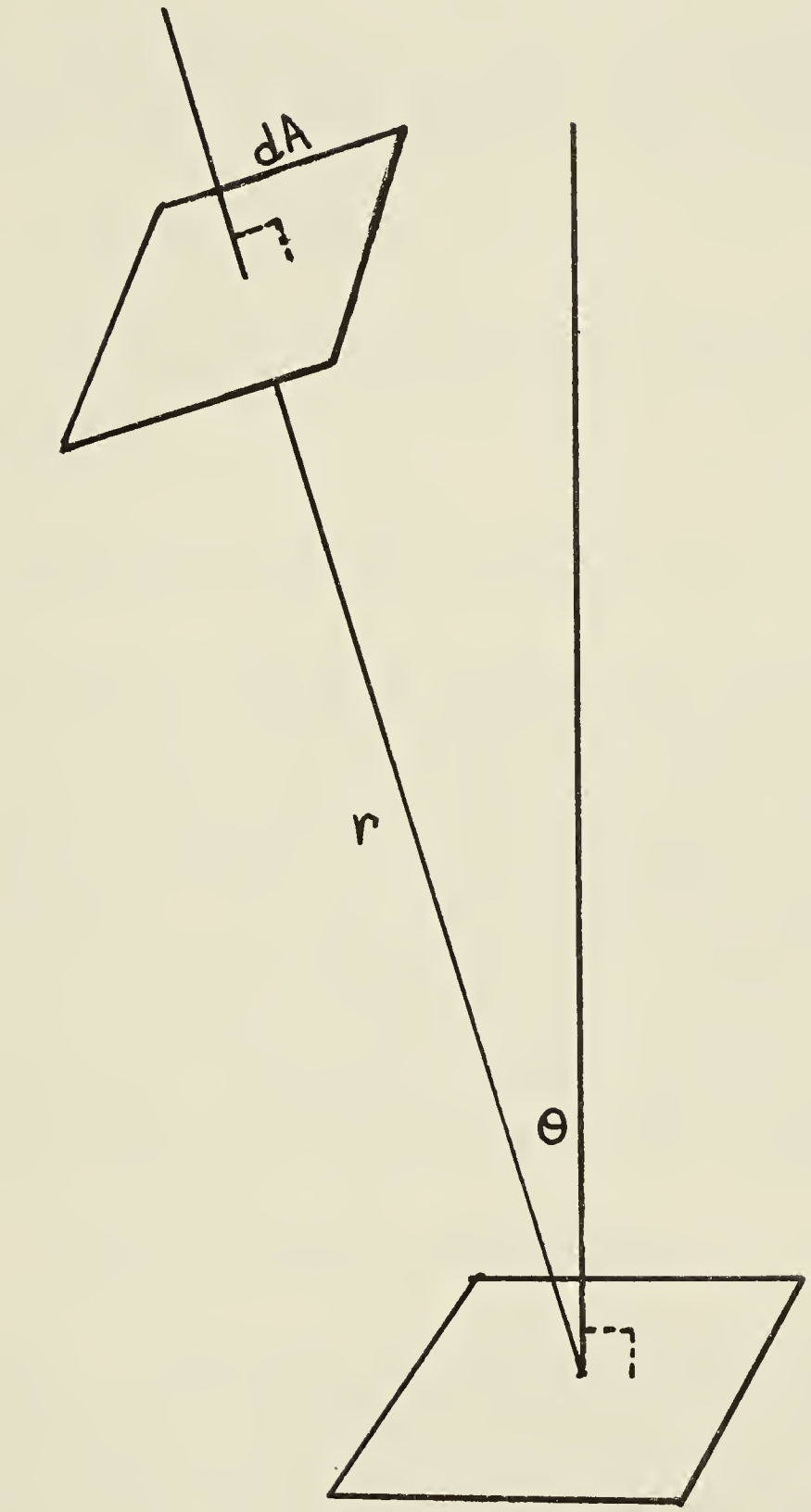
$$di = e\sqrt{\frac{2\varepsilon}{m}}\, n(\varepsilon)\ d\varepsilon \quad .$$

If i is the total current, then by a straightforward calculation it can be shown that the fractional current with energy between ε and $\varepsilon+d\varepsilon$ is

$$\frac{di}{i} = \frac{\varepsilon}{k_B T}\, e^{-\varepsilon/k_B T}\ d\varepsilon/k_B T \quad . \qquad \text{(A.2)}$$

Let us suppose that we have an area A emitting radiation into a half plane and a perpendicular area dA at radius r and angle θ to the perpendicular (see Figure A.2). Let dI be the total current hitting dA and I_o be the total current density emitted. Then the effective total area of A is $A\cos\theta$. This gives us Lamberts law

Figure A.2. Effective area of dA.



$$dI = \frac{C I_o A \cos\theta \, dA}{r^2}$$

where C is a constant. A short calculation will reveal that the total current emitted into a solid angle α is

$$i(\alpha) = I_o A \sin^2\alpha \quad . \qquad (A.3)$$

Let us define the index of refraction of an electron as

$$n(x,y,z) = C\sqrt{E - V(x,y,z)}$$

where n is the index of refraction, C is a constant we use to normalize n, E is the free space energy of the electron and V(x,y,z) is the potential that the electrons move in. This is the definition used by Langmuir (1937). It can be redefined to include a relativistic correction or a magnetic field (Klemperer, 1971). Let us also assume that Abbé's sine law holds. Then

$$n_o y_o \sin\theta_o = n_i y_i \sin\theta_i \qquad (A.4)$$

where n_o, n_i are the respective indices of refraction for the object and image positions, y_o, y_i. The respective sizes and θ_o,θ_i the respective angles of the rays with respect to the axis. The validity of the law for electrons has been demonstrated (Klemperer, 1971). We can also define the brightness, B_o, as the number of electrons emitted per second per unit solid angle

normal to A per unit area. If I_o is the current density of A, then

$$\int_0^{2\pi} \int_0^{\pi/2} \cos\theta \, d\Omega = I_o$$

$$I_o = \pi B_o \quad . \tag{A.5}$$

If we combine equations (A.2) and (A.3), we find that

$$\frac{di(\theta_o)}{A} = I_o \, \varepsilon/k_B T \, e^{-\varepsilon/k_B T} \, (d\,\varepsilon/k_B T) \, \sin^2 \theta_o$$

where $di(\theta_o)$ is the current of electrons with energy between ε and $\varepsilon+d\varepsilon$ emitted into half angle θ_o. If we substitute with (A.4) and (A.5) we find that

$$\frac{di(\theta_o)}{A} \frac{y_o^2}{y_i^2} = \pi B_o \frac{n_i^2}{n_o^2} \varepsilon/k_B T \, e^{-\varepsilon/k_B T} \, (d\,\varepsilon/k_B T) \, \sin^2 \theta_i \, .$$

Let

$$\frac{di(\theta_o)}{A} \frac{y_o^2}{y_i^2} = dI(\theta_o) \, .$$

Then $di(\theta_i)$ is the current density in the image due to electrons emitted with energies between ε and $\varepsilon+d\varepsilon$ in a solid angle θ_o. Note that $A(y_i^2/y_o^2)$ is the imaged size of area A. Then the equation becomes

$$dI(\theta_o) = \pi B_o \frac{n_i^2}{n_o^2} \varepsilon/k_B T \, e^{-\varepsilon/k_B T} \, (d\,\varepsilon/k_B T) \, \sin^2 \theta_i \, . \tag{A.6}$$

If we use an aperture in the imaging system such that the half angle of the imaged electrons is not greater than β, then there is a certain energy E_c below which all the electrons that are emitted will pass through the aperture. If an electron is emitted with an energy greater than E_c it will pass through the aperture only if it is emitted in a direction sufficiently close to the axis. At energy E_c an electron leaving at an angle $\pi/2$ with respect to the axis will just go through the aperture. Then using equation (A.4)

$$\sin^2\beta = \frac{n_o^2}{n_i^2}\frac{y_o^2}{y_i^2} = \frac{E_c}{E_2+E_c}\frac{1}{M^2} \qquad (A.7)$$

where E_2 is the energy that would be acquired by an electron emitted with no energy and M is the magnification. Rearranging gives

$$E_c = E_2\frac{M^2\sin^2\beta}{1-M^2\sin^2\beta} \quad . \qquad (A.8)$$

For $E < E_c$

$$\sin^2\theta_i = \frac{E}{E+E_c}\frac{1}{M_c^2} \quad . \qquad (A.9)$$

Then using (A.6) and (A.9) and the definition of the index of refraction we obtain for the current density at the image for $E < E_c$

$$dI\left(\frac{\pi}{2}\right) = \frac{\pi B_o}{M^2} \frac{E}{k_B T} e^{-E/k_B T} \, d(E/k_B T) \quad .$$

For $E > E_c$, $\sin^2\theta_i$ becomes $\sin^2\beta$ and (A.6) becomes

$$dI_o\left(\frac{\pi}{2}\right) = \pi B_o \left(\frac{E+E_2}{k_B T}\right) e^{-E/k_B T} \, d(E/k_B T) \sin^2\beta \quad .$$

Therefore

$$I = \frac{\pi B_o}{M^2} \int_0^{E_c} E/k_B T \, e^{-E/k_B T} \, d(E/k_B T)$$

$$+ \pi B_o \sin^2\beta \int_{E_c}^{\infty} \frac{E+E_2}{k_B T} e^{-E/k_B T} \, d(E/k_B T) \quad .$$

Then working the integrals and using (A.7) and (A.8) gives

$$I = \frac{\pi B_o}{M^2} \left[1 - (1 - M^2 \sin^2\beta) \, e^{-\frac{E_2 M^2 \sin^2\beta}{k_B T (1 - M^2 \sin^2\beta)}}\right] \quad .$$

If M is small enough (this assumption is certainly justified on the instrument used in the experiment) we can approximate the above relationship by

$$I = \pi B_o \left(\frac{E_2}{k_B T} + 1\right) \sin^2\beta \quad .$$

Then using (A.1) and (A.5) and assuming that β is small ($<15°$) we obtain

$$I = \frac{4\pi m_e e k_B^2 T^2}{h^3} \left(\frac{E_2}{k_B T} + 1\right) B^2 e^{-\phi/k_B T} \quad .$$

Then the brightness B of the gun is

$$B = \frac{I}{\beta^2} = \frac{4 m_e e k_B^2 T^2}{h^3} \left(\frac{E_2}{k_B T} + 1\right) e^{-\phi/k_B T} \quad .$$

If for example we use tungsten at 2700°K then ϕ = 4.52 eV, e = 1.602×10^{-19} coulombs, m_e = 9.11×10^{-28} gm, h = 6.626 10^{-21} erg-sec and k_B = 1.38×10^{-16} erg°K^{-1}. If E = 100,000 eV then

$$B' = \frac{I}{\pi\beta^2} = 4.7 \times 10^5 \text{ amp/cm}^2 \quad .$$

The above calculation is not in agreement with experiment (Cusack, 1967). The current density used to calculate it was

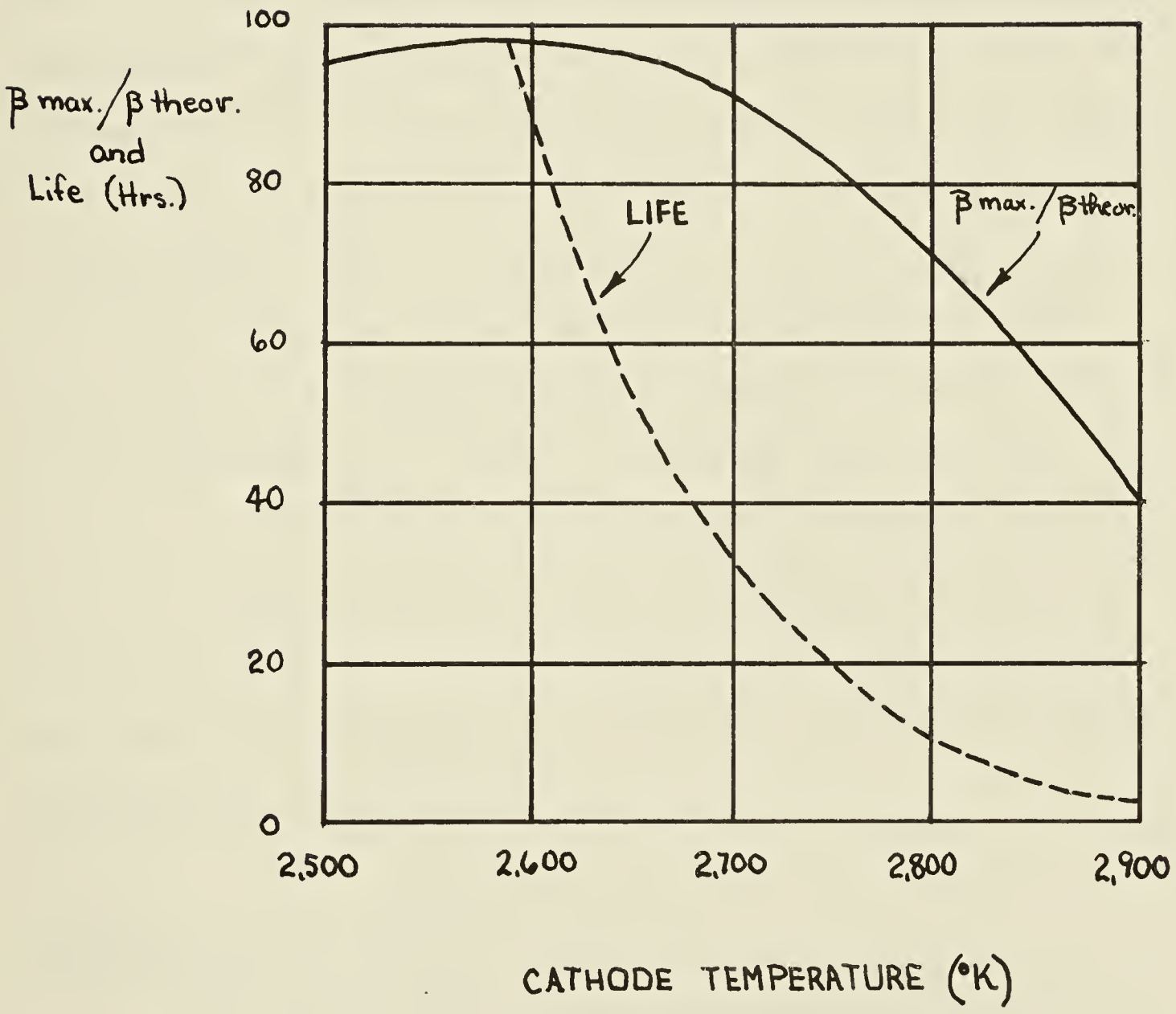
$$I = 120 T^2 e^{-\phi/k_B T} \quad .$$

But in practice, the constant is usually much smaller than 120. This is due in part to the fact that there is a small potential hump at the surface of any metal which low energy electrons must tunnel through in order to escape. This results in a reduction in the number of electrons escaping with low velocities and deviation from the Maxwell-Boltzmann distribution. If we calculate the brightness from the current density

of the cathode we get good agreement with experiment up to a point. In our calculations we have neglected the space charge effect. This results in a reduction in the brightness below the theoretical value at higher temperatures. In addition filament life decreased rapidly with temperature past a certain value. A graph of these effects, due to Haine (1961), is shown in Figure A.3. Note that in this case there is an optimum temperature of about twenty-seven hundred degrees Kelvin that gives us nearly the maximum brightness and also results us reasonable filament life.

The expression we derived above for the current density varies rapidly with temperature. This could result in small changes in the temperature of the filament giving rise to undesirable fluctuations in the current. This is normally avoided by employing a feedback process. In addition to this we need a focusing action to collect the electrons. We can do this by employing a bias voltage on the Wehnelt cylinder. This affects the shape of the field in the vicinity of the cathode. If the bias is too strong and accelerating field will not penetrate as far as the cathode and the beam current will be reduced. If it is too weak then the focusing action is lost and the brightness is once again diminished. It is difficult to provide an

Figure A.3. Life and brightness of filaments.

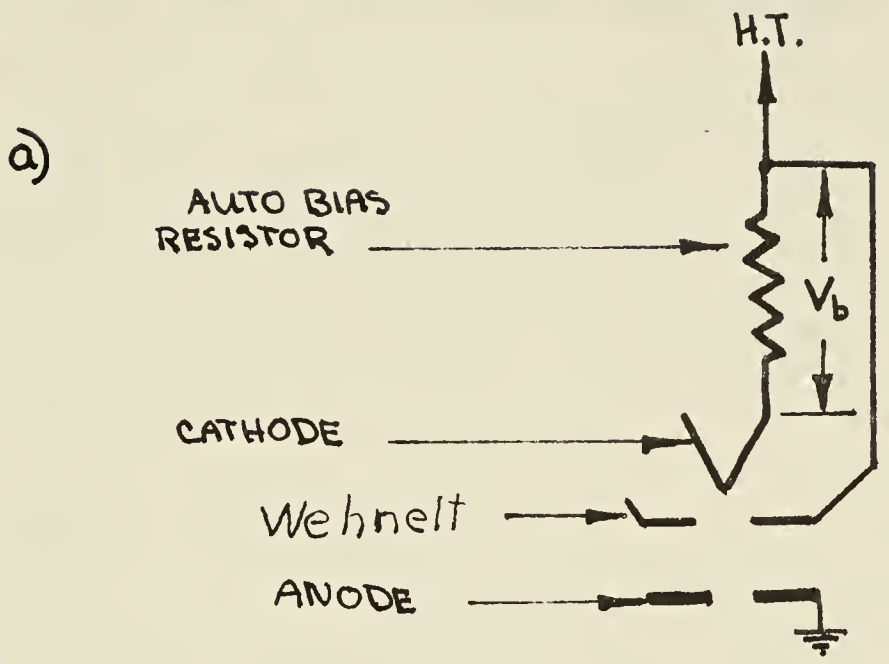


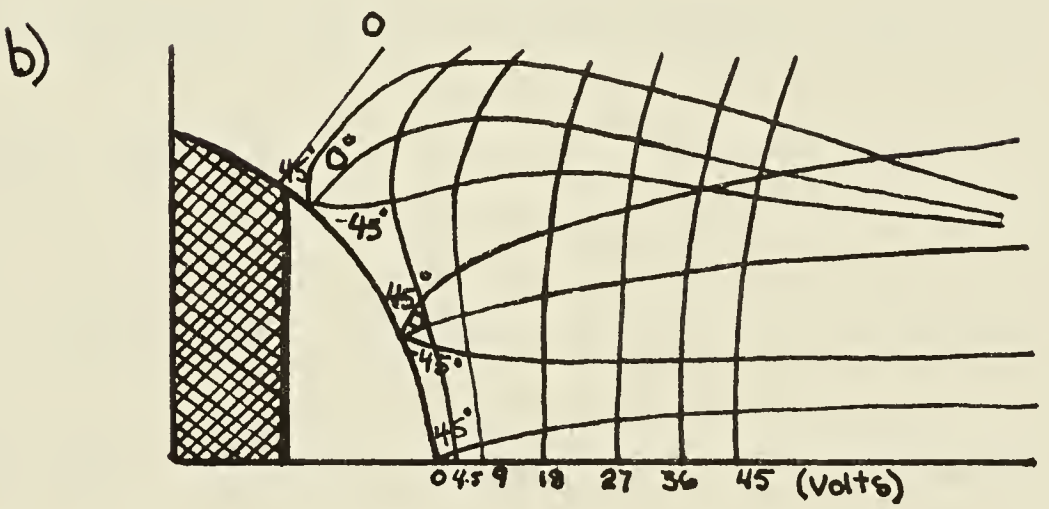
external bias voltage to a filament at a hundred thousand volts and therefore auto-bias is usually used. This has an additional advantage - the auto-bias resistor can stabilize the gun near the optimum conditions. If the current increases the bias voltage increases, decreasing the current. If the current decreases the bias voltage decreases, increasing the current. Thus the gun is stable through negative feedback and we need only choose one optimum value for the resistor. Figure A.4 illustrates this. Part a) shows the auto-bias system, part b) shows too low a bias and part d) too high a bias. Notice how the electrons will not come to a common focus in either case b) or case d) but will under optimum bias in part c). All these illustrations are due to Haine (1961). Haine also found that increasing the filament height increases the total beam current available but not the brightness implying that the gun crossover spot size also increased.

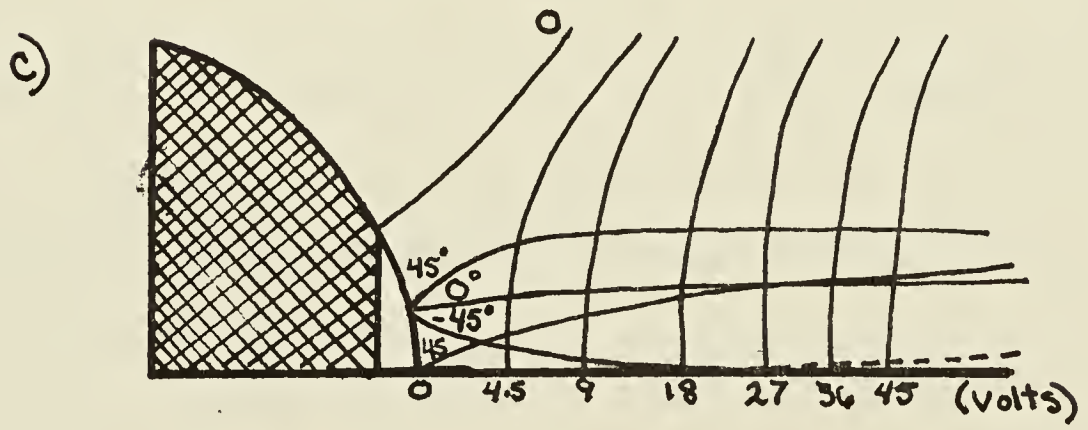
Section 2: The Lenses

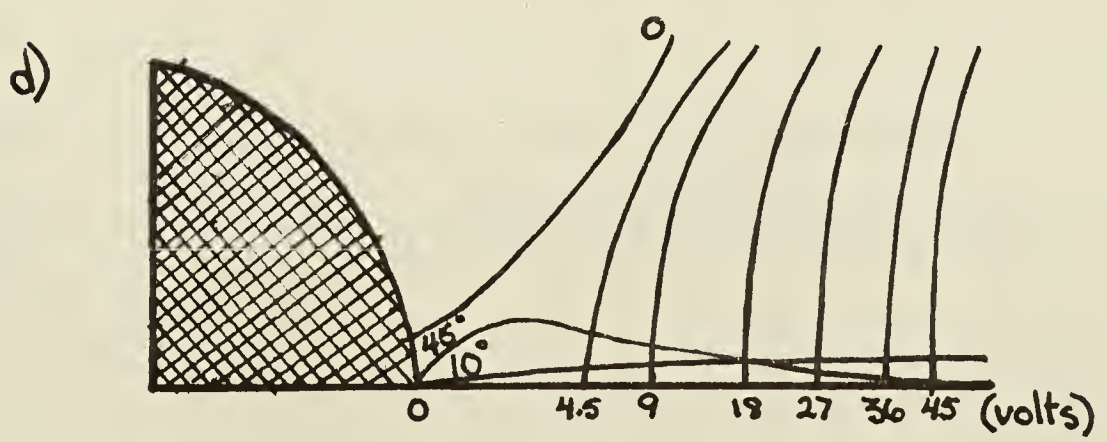
The lenses have a large influence on the resolution available from a STEM instrument. In this section we shall consider the paraxial ray approximation in order to illustrate the behavior of electrons in a magnetic focusing system. Then we shall see how the

Figure A.4. Effect of biasing.









higher order terms that were ignored in this approximation give rise to spherical aberration. We shall also see how chromatic aberration arises. The treatment that we shall follow is from El-Kareh and El-Kareh.

The equation of motion for an electron in a magnetic field is

$$m \frac{d^2\vec{r}}{dt^2} = - e(\vec{v} \times \vec{B}) \quad .$$

If we use cylindrical coordinates (r,ϕ,z) and assume rotational symmetry $(B_\phi = 0)$ we can rewrite this as

$$\ddot{z} = \frac{e}{m} \; r\dot{\phi}B_r$$

$$\ddot{r} = - \frac{e}{m} \; r\dot{\phi}B_z + r\dot{\phi}^2 \qquad \text{(A.10)}$$

$$\frac{d}{dt} (r^2\dot{\phi}) = \frac{e}{m} (r\dot{r}B_z - \dot{z}rB_r) \qquad \text{(A.11)}$$

Now using Maxwell's equations with $E = 0$ and assuming free space (U constant), it is possible by doing a series expansion on the magnetic scalar potential to show that to first order

$$B_r = - \frac{r}{2} B_z \quad .$$

It turns out that the error in this approximation increases as r increases. Inserting this relationship in (A.11) gives

$$\frac{d}{dt}\,(r^2\dot{\phi}) = \frac{e}{m}\,(B_z r\dot{r} - \frac{r^2}{2}\,\dot{z}\,\frac{dB_z}{dz})$$

$$= \frac{e}{m}\,\frac{d}{dt}\,(\frac{1}{2}\,B_z r^2) \quad .$$

Integrating gives

$$mr^2\dot{\phi} = \frac{e}{2}\,B_2 r^2 + C$$

where C is a constant. In order to simplify the calculation, we will assume that outside the lens $\dot{\phi} = 0$, i.e. that we are dealing only with meridonal rays, then it is easy to show that C = 0. Then

$$\dot{\phi} = \frac{e}{2m}\,B_z \quad .$$

If we substitute this in (A.10) we get

$$\ddot{r} = -\frac{e}{m}\,r\frac{e}{2m}\,B_z B_r + r(\frac{e}{2m}\,B_z)^2$$

$$= -(\frac{e}{2m})^2\,r\,B_z^2 \quad .$$

If we use the classical relationship

$$v_z = \frac{dz}{dt} = (\frac{2eV}{m})^{\frac{1}{2}}$$

we obtain

$$r'' = -(\frac{e}{8mV})\,B_z^2 r$$

where the prime indicates differentiation with respect to z. If the accelerating voltage is high it is necessary to use the relativistically corrected voltage

in place of the actual voltage but the relationship still holds. Notice that the negative sign means that the force is always directed towards the axis and that the force is proportional to the distance from the axis. This is the principal of a focussing system.

So far we have carried out our computations using the paraxial ray approximation. However, just as in light optics, the off axis rays give rise to aberrations. To calculate these let us consider a cartesian coordinate system in which the magnetic field is symmetric around the z axis. Then

$$\frac{1}{2} m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) = e\phi$$

$$m\ddot{x} = -e\dot{y}B_z + e\dot{z}B_y = -e\dot{y}B_z + e\left(\frac{y}{r}\right)\dot{z}B_r$$

$$m\ddot{y} = -e\dot{z}B_x + e\dot{x}B_z = e\dot{x}B_z - e\left(\frac{x}{r}\right)\dot{z}B_r \quad .$$

Let

$$w = x + jy, \quad \bar{w} = x - jy, \quad j = \sqrt{-1} \quad .$$

This gives us (the prime indicates differentiation by z)

$$\frac{1}{2} m\dot{z}^2(1 + w'\bar{w}') = e\phi$$

$$m\dot{z}\,\frac{d}{dz}(w'\dot{z}) = je\dot{z}\left(w'B_z - \frac{w}{r}B_r\right) \quad .$$

Therefore

$$\frac{2\phi^{\frac{1}{2}}}{(1+w_1'\bar{w}')^{\frac{1}{2}}}\frac{d}{dz}\frac{w'\phi^{\frac{1}{2}}}{(1+w'\bar{w}')^{\frac{1}{2}}} = j\left(\frac{e}{m}\right)^{\frac{1}{2}}\left(\frac{2\phi}{1+w'\bar{w}'}\right)^{\frac{1}{2}}(w'B_z - \frac{w}{r}B_r) \quad .$$

Let

$$u = we^{-jk} \quad , \quad \bar{u} = \bar{w}e^{jk} \quad , \quad k = \left(\frac{e}{8m}\right)^{\frac{1}{2}}\int_{z_o}^{z}\frac{B}{\phi^{\frac{1}{2}}}\,dz \quad .$$

If we assume that $|w'\bar{w}'|<1$ for the purpose of series expansion and do a great deal of math that it is not profitable to go into here we obtain

$$u'' + \frac{eB^2}{8m}u = W$$

where W contains all the higher order terms in u. To third order

$$W = (u'\bar{u}' + \frac{eB^2}{8m}u\bar{u} + j\left(\frac{e}{8m}\right)^{\frac{1}{2}}B(uu' - \bar{u}'u))\cdot$$

$$\cdot(u'' - \frac{eB^2u}{8m} + j\left(\frac{e}{8m}\right)^{\frac{1}{2}}(2Bu' + B'u)$$

$$\frac{\phi}{2}\left(\frac{eBB'}{4m\phi} + j\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}B'(u\bar{u}' - u'\bar{u})\right)\cdot$$

$$\cdot(u' + jB\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}u) - \frac{j}{2}\left(\frac{2e\phi}{m}\right)^{\frac{1}{2}}[B''\frac{u\bar{u}}{4}(u' + u\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}Bu) + \frac{B'''}{16}u^2\bar{u})]$$

$$- \frac{j}{4}\left(\frac{2e\phi}{m}\right)^{\frac{1}{2}}(u'\bar{u}' + \frac{eB^2}{8m\phi}u\bar{u} + j\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}B(u\bar{u} - u'\bar{u}))\cdot$$

$$\cdot(Bu' + j\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}B^2u + \frac{B'}{2}u) \quad .$$

We can define the two solutions to the paraxial equation

$$u_p'' + \frac{eB^2}{8m} u_p = 0$$

as

$$r_o \quad \text{when} \quad r_o(z_o) = 0 \quad r'(z_o) = 1$$

$$r_a \quad \text{when} \quad r_a(z_o) = 1 \quad r(z_a) = 0$$

where the subscript 0 represents the object position and a the aperture position. Then

$$u = \frac{u_a r_o}{r_{oa}} + u_o r_a \quad , \quad \bar{u} = \frac{\bar{u}_a r_o}{r_{oa}} + \bar{u}_o r_a$$

where u_a is the position of the ray in the aperture plane, u_o the position of the ray in the object plane and r_{oa} the position of the first solution in the aperture plane. Let us assume a solution of the form

$$u = u_p + \Delta u \quad .$$

Let us also assume that Δu is a function of r_o and r_a and that Δu is small enough to be ignored on the right side. Let us call W after u paraxial is substituted W'. After some math it is possible to show that in the image plane

$$\Delta u = \frac{M}{\phi} \int_{z_o}^{z_i} r_o W' dz$$

$$= \frac{M}{\phi}\int_{z_o}^{z_i} r_o\left(\frac{eB^2}{8m\phi}\frac{r_o^2}{r_{oa}}+\frac{r_o'^2}{r_{oa}}\right)\cdot\left[\left(-\frac{eB^2}{8m\phi}+j\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}\frac{B'\phi}{2}\right)\frac{r_o}{r_{oa}}\right.$$

$$\left.+\ j\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}B\phi\frac{r_o'}{r_{oa}}\right]u_a^2\bar{u}_a$$

$$+\ c_2u_a^2\bar{u}_o+c_3u_au_o\bar{u}_a+c_4u_au_o\bar{u}_o+c_5u_au_o\bar{u}_a$$

$$+\ c_6u_au_o\bar{u}_o+c_7u_o^2\bar{u}_a+c_8u_6^2\bar{u}_6+\frac{eBB'}{8m}\left(jB\left(\frac{e}{8m\phi}\right)^{\frac{1}{2}}\frac{r_o}{r_{oa}}+\frac{r_o''}{r_{oa}}\right)u_a^2\bar{u}$$

$$+\ c_9u_au_o\bar{u}_a+c_{10}u_a^2\bar{u}_o+c_{11}u_au_ou_o+c_{12}u_au_o\bar{u}_a$$

$$+\ c_{13}u_6^2\bar{u}_a+c_{15}u_au_o\bar{u}_o+c_{16}u_6^2\bar{u}_o+\left[\frac{eBB''}{16m}-j\left(\frac{e\phi}{2m}\right)^{\frac{1}{2}}\frac{B''}{16}\frac{r_o^3}{r_{oa}^2}\right.$$

$$\left.-\ j\left(\frac{e\phi}{2m}\right)^{\frac{1}{2}}\frac{B''}{16}\frac{r_or_o}{3}\right]u_a^2\bar{u}_a$$

$$+\ c_{17}u_au_o\bar{u}+c_{13}u_o^2u_a+c_{19}u_a^2\bar{u}_o+c_{20}u_au_o\bar{u}_o+c_{21}u_o^2\bar{u}_o]dz\ .$$

Here the c's indicate some function of the magnetic field, the various r's, M and the accelerating voltage. Only the term that has been written out in full cannot be made zero by considering a ray starting on the axis at the object. The spherical aberration is the amount that a point object on the axis expands to in the image. Therefore this is the spherical aberration. After rearrangement we get the spherical aberration coefficient

$$c_1 = - \frac{M}{16 r_{ao}^3} \phi^{-2} \int_{z_o}^{z_i} \left(\frac{e\phi B'^2}{m} + \frac{3e^4 B^4}{8m^2} - \frac{e\phi B^2 r_o'^2}{m r_{oa}^2} \right) r^4 dz$$

$$= - \frac{M}{16 r_{ao}^3} \phi \int_{z_o}^{z_i} \left\{ \frac{e}{m\phi} \left(B' + \frac{B r_o'}{r_o} \right)^2 + \frac{eB^2}{m\phi} \frac{r_o'^2}{r_o^2} + \left(\frac{eB^2}{2m\phi} \right)^2 \right\} dz \ .$$

The last expression is intended to show that c_1 is a sum of squares and is therefore always negative. The best we can hope to do therefore is to minimize it. Thus we can write that the radius of the image of u point is

$$r_i = c_1 u_a^2 \bar{u}_a \ .$$

If we consider the aperture as located in the principal plane, then

$$|u_a| - f_o \tan \theta, \qquad \tan \theta \simeq \theta$$

$$|\Delta r_i| = c_1 f_o^3 \theta^3 = - c \theta_3 \ .$$

However, it turns out that the Gaussian plane is not the plane with the disk of minimum confusion. Let e be an extreme ray and F_e its focal point, m be a ray between the extreme ray and the axis and F_o the axial focus. Figure A.5 is a diagram of this situation. Then the radius of the disk of minimum confusion between the two rays is

$$h = F_e B \tan \theta_e = F_m B \tan \theta_m$$

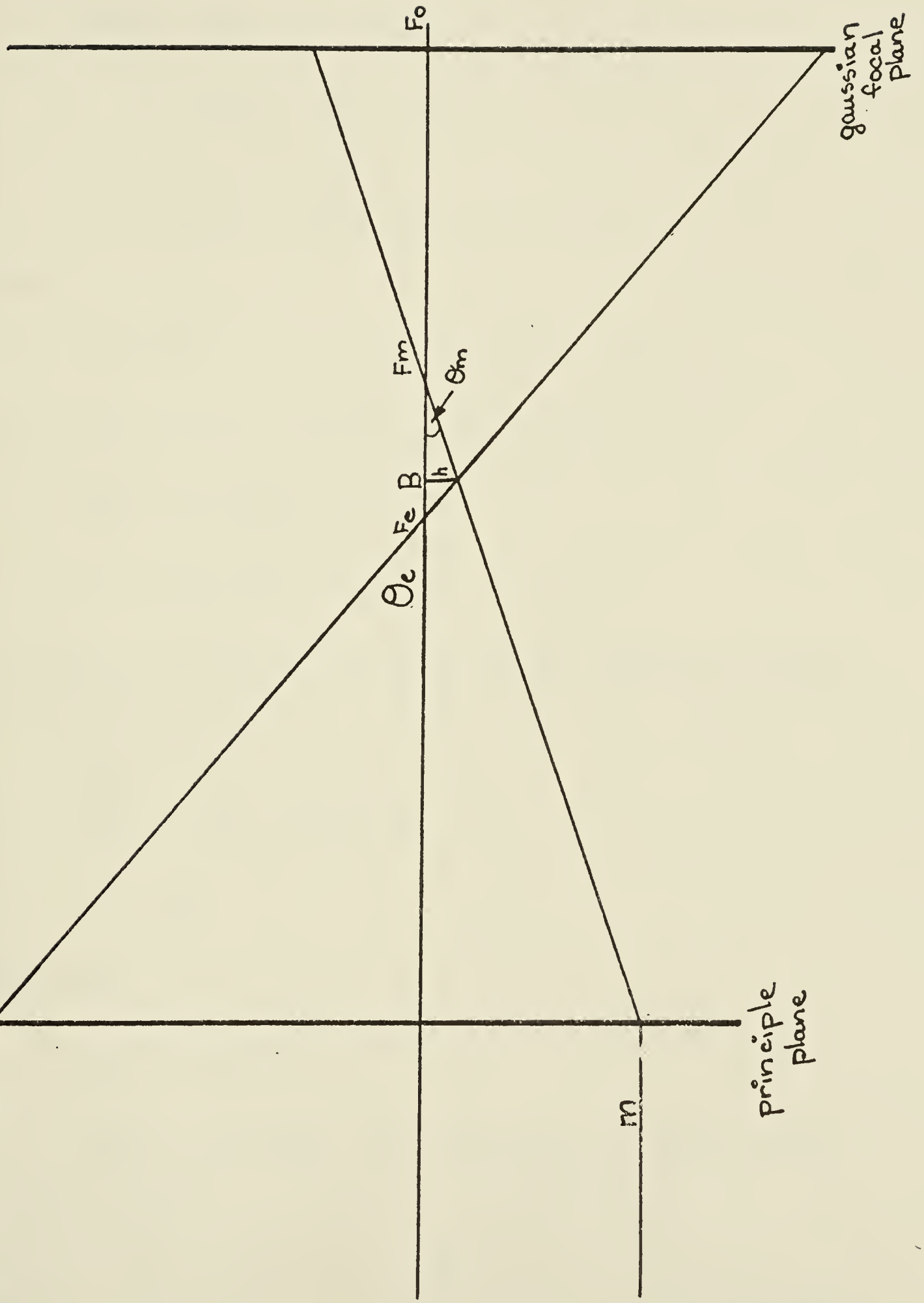


Figure A.5. Effect of spherical aberration.

but

$$F_m B = F_e F_o - F_e B - F_m F_o$$

$$= c_s \theta_e^2 - c_s \theta_m^2 - F_e B$$

$$h = (c_s \theta_e^2 - c_s \theta_m^2 - F_e B) \tan \theta_m \quad .$$

Then

$$F_e B(\tan \theta_e + \tan \theta_m) = c_s(\theta_e^2 - \theta_m^2)\tan \theta_m \quad .$$

If we let $\tan \theta = \theta$ (small angle approximation) we get

$$F_e B = c_s \theta_m (\theta_e - \theta_m)$$

$$h = c_s \theta_e \theta_m (\theta_e - \theta_m) \quad .$$

To find the maximum value of h we set the first derivative equal to zero

$$\frac{dh}{d\theta_m} = c_s \theta_e^2 - 2c_s \theta_e \theta_m = 0$$

$$\theta_m = \theta_e/2 \quad .$$

Therefore

$$h_{max} = c_s \theta_e^3/4 \quad .$$

There is another lens aberration that we shall have to consider. If there is a variation in the magnetic field (caused by a variation in the lens current) or a variation in the energy of the electrons

(caused for instance by the thermal spread) then chromatic aberration arises. We can write down the paraxial ray equation

$$r'' + \frac{e}{8mV} B_z^2 r = 0 \quad .$$

Let us assume that V becomes V + ΔV and B becomes B + ΔB where $\Delta B = B(\Delta B_{max}/B_{max})$. Then leaving out the second order terms

$$(V + \Delta V)r'' + \frac{e}{8m}(B_z^2 + 2B_z^2 \frac{\Delta B_{max}}{B_{max}}) = 0 \quad .$$

Now we will assume that p(z) is the solution of the paraxial equation such that it leaves the object at r = 0 and arrives at the image at r = 0. Let q(z) be the solution that leaves the object at r = a and arrives at the image at r = Ma. We are interested in the amount that a point object deviates from a point at the image. Let us assume a solution to the above equation of the form

$$r(z) = p(z) + \varepsilon(z)$$

where $\varepsilon(z)$ is the variation due to the perturbation. Substitution yields $\varepsilon'' + \frac{e}{8mV} B_z^2 = -\frac{\Delta V}{V} p'' - \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{B_{max}} p$ where we have kept only first order terms and made use of the paraxial equation to eliminate two terms. The solution of this equation is the solution to the homogeneous part $(c_1 p(z) + c_2 q(z))$ plus a particular solution.

It is reasonalbe that the particular solution can be written $m_1(z)p(z)$. Then

$$\varepsilon(z) = (c_1 + m_1(z))p(z) + c_2 q(a) \quad .$$

We could have absorbed the c_1 in $m(z)$. However, we will instead choose as a solution

$$\varepsilon(z) = G_1(z)p(z) + G_2(z)q(z) \quad . \tag{A.12}$$

We know from the theory of differential equations that because we did not choose $G_2(z)$ as a constant, we can choose a relation between $G_1(z)$ and $G_2(z)$. We choose

$$G_1'p + G'q = 0 \quad . \tag{A.13}$$

Then

$$\varepsilon'(z) = G_1p' + G_2q'$$

$$\varepsilon'' = G_1'p + G_1p'' + G_2'q' + G_2q'' \quad .$$

Substitution gives

$$G_1'p' + G_1p'' + G_2'q' + G_2q''$$

$$\frac{e}{8mV}(G_1p + G_2q) = -\frac{\Delta V}{V}p'' - \frac{e}{4m}\frac{B^2}{V}\frac{\Delta B_{max}}{B_{max}}p \quad .$$

Using the paraxial solution gives

$$G_1'p' + G_2'q' = -\frac{\Delta V}{V}p'' - \frac{e}{4m}\frac{B^2}{V}\frac{\Delta B_{max}}{B_{max}}p \quad .$$

Substitution with (2.12) gives

$$-G_2' p' \frac{q}{p} + G_2' q' = -\frac{\Delta V}{V} p'' - \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{\Delta B_{max}} p$$

$$G_2' = p(\frac{\Delta V}{V} p'' + \frac{eB^2}{4mV} \frac{\Delta B_{max}}{B_{max}} p)/p'q - q'p \ .$$

Similarly

$$G_1' p - G_1' q' \frac{p}{q} = -\frac{\Delta V}{V} p'' - \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{B_{max}} p$$

$$G_1' = q(\frac{\Delta V}{V} p'' - \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{B_{max}} p)/pq' - qp' \ .$$

Integrating and using (A.12) gives

$$\varepsilon = c_1 p + p \int_{x_o}^{z} \left\{ q(\frac{\Delta V}{V} p'' + \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{B_{max}} p)/(pq' - qp') \right\} dz$$

$$+ q \int_{z_o}^{z} \left\{ p(\frac{\Delta V}{V} p'' + \frac{e}{4m} \frac{B^2}{V} \frac{\Delta B_{max}}{B_{max}} p)/(p'q - q'p) \right\} dz$$

where z_o is the object plane. We know by assumption that

$$\varepsilon(z_o) = 0$$

but

$$\varepsilon(z_o) = c_1 p(z_o) + c_2 q(z_o) + c_2 q(z_o) \ .$$

Therefore $c_2 = 0$. We also know that

$$\varepsilon'(z_o) = 0 \ .$$

Because

$$\varepsilon'(z_o) = c_1 p'(z)$$

we know

$$c_1 = 0 \ .$$

Since p and q are solutions

$$p'' + \frac{e}{8mV} B_z^2 p = 0$$

$$q'' + \frac{e}{8mV} B_z^2 q = 0 \ .$$

If we multiply a) by q and b) by p and subtract we get

$$p''q - p''p = 0$$

$$\frac{d}{dz} (p'q - q'p) = 0 \ .$$

Integrating gives

$$p'q - q'p = c_3 \ .$$

Now we know that $p(z)$ is zero at $z = z_i$ where z_i is the image plane. Then

$$c_3 = p'(z_i)q(z_i) \ .$$

This gives us

$$\varepsilon(z_i) = q(z_i) \int_{z_o}^{z_i} \frac{\frac{1}{V} p(\Delta V p'' + \frac{eB^2}{4m} \frac{\Delta B_{max}}{B_{max}} p) dz}{p'(z_i)q(z_i)}$$

$$= \frac{1}{Vp'(z_i)} \int_{z_o}^{z_i} p(\Delta V p'' + \frac{eB^2}{4m} \frac{\Delta B_{max}}{B_{max}} p) dz \ .$$

Now if we substitute the paraxial equation we get

$$\varepsilon(z_i) = \frac{1}{Vp'(z_i)} \int_{z_o}^{z_i} p(\Delta V p'' - \frac{2V\Delta B_{max}}{B_{max}} p'')dz$$

$$= \frac{1}{p'(z_i)} (\frac{\Delta V}{V} - \frac{2\Delta B_{max}}{B_{max}}) \int_{z_o}^{z_i} p''(z)^2 p(z)dz \ .$$

Integrating by parts gives

$$\varepsilon(z_i) = - \frac{1}{p'(z_i)} (\frac{\Delta V}{V} - \frac{2\Delta B_{max}}{B_{max}}) \int_{z_o}^{z_i} p'^2 dz$$

$$= -p'(z_i) (\frac{\Delta V}{V} - \frac{2\Delta B_{max}}{B_{max}}) \int_{z_o}^{z_i} \frac{p'(z)^2}{p'(z_i)^2} \ .$$

The integral is just a constant. Let us call this c_c. Usually the effects due to fluctuations in the magnetic field are smaller than those due to the energy spread of the electrons leaving the cathode. Now

$$p'(z_i) = \tan \theta_i \simeq \theta_i \ .$$

Therefore

$$|r_i| = c_c \frac{\Delta V}{V} \theta_i$$

where r_i is the radius of the disk of confusion and θ_i the angle at the image. First it should be noted that we can always choose the focus so that we get our existing focus for the centre of the energy spread allows us to write

$$d = 2|r_i| = 2c_c \frac{\frac{1}{2}\Delta V}{V} \theta_i = c_c \frac{\Delta V}{V} \theta_i \ .$$

The two obvious contributions to the energy spread ΔV are gun instability (changes in the accelerating voltage) and the energy spread inherent in thermal emission. There is however one other contribution - the Boersh effect. Here electrons in the beam collide and some of their transverse momentum is converted to momentum parallel to the axis. This effect can increase the energy spread by 1 or 2 eV (Hawkes, 1972).

It was shown that there is a limit to the brightness of a thermionic gun of about 3×10^5 amps/cm^2/sterad. The outline of the proof that spherical aberration cannot be corrected was considered. Finally, an expression has been derived for the chromatic aberration coefficient.